THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 56, PART 4 1 July 1944 No. 316

THE LIQUID STATE

BY J. H. HILDEBRAND,
University of California

Twenty-eighth Guthrie Lecture: delivered in London 26 April 1944 and in Oxford 29 April 1944

§ 1. INTRODUCTION

THE invitation extended to me by the President and Council of this distinguished Society to deliver one of the lectures in memory of its founder, Professor Guthrie, has put a strain upon my modesty, relieved only by seeing in it something I prize even more highly than a scientific honour, namely, a kindly gesture from British scientists, with whom it has been my privilege to associate in our present common endeavour.

I selected for my topic *The Liquid State*, partly because I expected the subject at least to be of general interest, but more particularly because it is one about which I have most to say. My "terms of reference" permit me to confine my attention mainly to my own adventures in this field, or, perhaps I should say, my voyages upon this sea. I trust this limitation will not be misinterpreted as a claim to have been the principal voyager. A number of able scientists have recently been active in extending our knowledge of liquids, among whom may be mentioned Lennard-Jones, Devonshire, Fowler, Rushbrooke, Guggenheim, Eyring, Kirkwood, Rice, Huggins, Gingrich, Mayer and Scatchard. My own work is only a minor part of the total, so minor that I can outline it in one hour, which I could not do for the whole subject, particularly in the midst of my current pressing duties and so far from my own files.

There are two conventional approaches to the liquid state (Hildebrand, 1936 a, 1937 a), one via the gaseous state, the other via the solid; the one leading to the view that liquids are compressed gases, the other that they are disordered solids. Both approaches have been fruitful. The former is exemplified by the van der Waals equation, which is helpful qualitatively, despite its quantitative shortcomings. If slightly transformed, it may be compared, term for term, with a purely thermodynamic equation of state, as follows:

$$p + \frac{a}{\mathrm{v}^2} = T\frac{\boldsymbol{R}}{\mathrm{v}-b},$$

$$p + \left(\frac{\partial E}{\partial v}\right)_T = T\left(\frac{\partial p}{\partial T}\right)_v = T\left(\frac{\partial S}{\partial v}\right)_T.$$

where p denotes pressure; v, volume; v, molal volume; T, absolute temperature; E, energy; S, entropy; $\boldsymbol{R}$, the gas constant; and a and b the van der Waals "constants". The terms a/v^2 and $\boldsymbol{R}/(v-b)$ do not serve to reproduce their thermodynamic counterparts very accurately over any considerable range. My former co-workers, Dr. Westwater and Dr. Carter, measured $(\partial p/\partial T)_v$ for a

number of liquids with great care. There have been many proposals for corrected equations of state, but it is probably not unfair to say that the extra constants they introduce represent more success in fitting certain sets of experimental data over limited ranges than progress towards a satisfactory theoretical basis.

An interesting and ingenious picture of the liquid state has been suggested by Eyring, called the "Theory of Holes", which accounts for the *Law of Rectilinear Diameter* for the curves of liquid and gas density by assuming that a liquid is a sort of inverse of its own vapour, containing holes where the vapour has molecules.

The analogy between liquid and gas is clearly overworked in the theoretical uses of osmotic pressure. The formal identity between the ideal gas equation and the van't Hoff law for osmotic pressure makes it so easy for teachers of physical chemistry to derive therewith certain useful equations that whole generations of students have come to regard a solvent as merely providing space for pseudo-gaseous solute molecules. It is so simple that faith in the results has extended far beyond the limitations pointed out by van't Hoff himself. I was rebuked, in my youth, by an older scientist, for daring to attack the great van't Hoff when I had done little more than restate those same limitations. Now, in my later years, I am still so unrepentant as to insist that students would be better served by omitting all mention of osmotic pressure except as a comparatively minor corollary of Henry's law or Raoult's law. As justification I shall present results in the handling of concentrated solutions quite impossible to achieve by any treatment based upon osmotic pressure.

The solid state of matter has yielded its secrets under scrutiny by the aid of x rays, supplemented by measurements of specific heat, etc., and we now possess accurate knowledge of both the structures and thermodynamic properties of crystals. This has invited an approach to the liquid state from that direction, picturing a liquid as a solid lattice somewhat disturbed by the process of melting. This is unquestionably a legitimate approach, and it has yielded valuable results. Most scientific progress consists in making closer approximations, and a fruitful method is not to be unduly deprecated by reason of lack of finality.

§ 2. MOLECULAR DISTRIBUTION

I invite you, however, to follow me now along a rather different path. Instead of approaching the liquid state via the gaseous or solid states, let us try a direct, experimental approach by the aid of x rays. The scatter pattern of a liquid does not suggest a crystal lattice at all. It can be analysed, as shown by Debye and Menke (Gingrich, 1943), to give an average number of molecular centres at various distances from a central molecule. Figure 1 B shows the result for liquid mercury, with the solid structure superimposed for comparison.

An alternative method of representing the same result is by means of a variable called the *distribution function*, W. This has the following meaning:—The number of molecular centres in a spherical shell of radius r and thickness dr is the number of molecular centres per c.c., $N\nu$, N denoting the Avogadro number and ν the molal volume, multiplied by the volume of the shell, $4\pi r^2 dr$. This is true, however, only when r is large, for when it is sufficiently small, a correcting

factor, W, must be introduced to represent the structure resulting from the spacial requirements of the molecules. With spherical molecules, considered for the moment as incompressible, no surrounding molecule can approach the central one more closely than twice the molecular radius, hence $W=0$ at distances

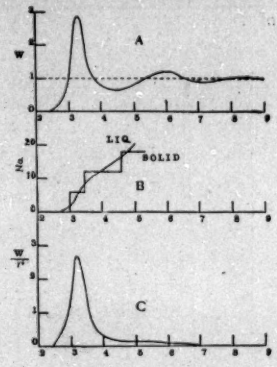
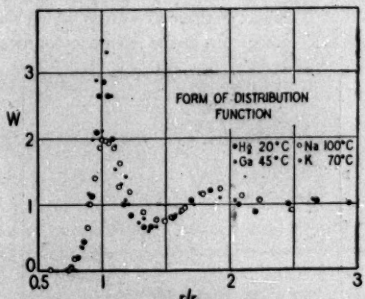


Figure 1.

less than $2r$. At a slightly greater distance, however, a shell has more than unit probability of enclosing centres, so that here $W>1$. It must fall below unity again between the first and second surrounding layers, and so on, oscillating ever less, above and below unity, as the partial order of the first shell is obscured



Form of distribution function.

Figure 2.

at greater distances by thermal agitation. The plot of W against r for mercury is shown in figure 1 A.

This kind of order is only short-range and is a purely geometric affair, such as might be obtained by vigorously shaking tennis balls in a box a little bigger than the space required for close-packing. We may compare molecules of

different size by plotting W against r/r_m, r_m being the position of the first maximum. We see from figure 2 that the curves for the four metals, Hg, Ga, Na and K (the latter two by Gingrich, Trimble and Thomas, 1943), when plotted in this way, all superimpose quite nicely.

Dr. M. E. Morrell (1936 b), in my laboratory, constructed a model of such a liquid by putting a lot of gelatine spheres, including a few coloured ones, in a

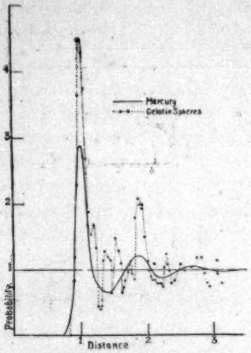


Figure 3.

cubic vessel containing a boiled gelatine solution of the same density and refractive index as the spheres, and in which, therefore, all disappeared from view except the coloured ones. The vessel was shaken and a spark photograph immediately taken by aid of a mirror in two directions, permitting the position of each coloured ball to be determined with respect to three axes from which their radial distances from each other were calculated. From the frequency of these distances, values

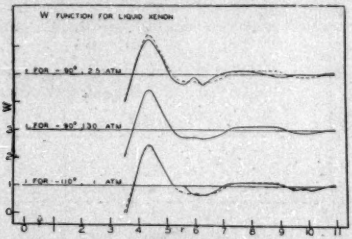


Figure 4.

of W were calculated, shown in figure 3 superimposed upon the curve for mercury. The higher peak merely indicates closer crowding, corresponding to lower temperature. This peak flattened progressively as higher temperatures were simulated by greater dilution.

Another of my collaborators, Dr. J. A. Campbell (1943 b), has recently determined the scattering for xenon at 2·5 atmos. and $-90°$, at 130 atmos. and $-90°$, and at 1 atmo. and $-110°$, shown in figure 4; the second and third

conditions were chosen to give the same liquid density. We wished to find out whether W is a pure volume function, as one might expect it to be, except for slightly smaller minimum distances at higher temperatures, due to greater kinetic energy. Figure 4 shows that the two curves can be superimposed with but very minor deviations. Indeed, the curves for xenon and for argon, at nearly corresponding states (the latter obtained by Gingrich (1943) and Eisenstein and Gingrich (1942)) fit each other almost as well (figure 5).

These experimental results indicate that the form of W vs. r offers a significant description of the structure of liquids composed of spherical molecules, and further, that it is nearly identical for all such liquids at equal degrees of expansion. It is subject to fairly accurate experimental determination. I have been trying for some time to find a mathematician willing and able to calculate it, at least for the case of tennis balls, but thus far without success. Kirkwood and Boggs (1942) have made a notable approximation but have not quite equalled x rays. I propose, therefore, provisionally, that the mean experimental curves of

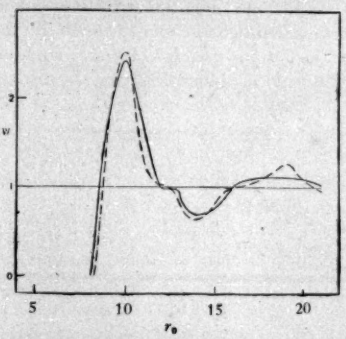


Figure 5.

W vs. r/r_0, at different temperatures, where r_0 is the intermolecular distance at 0° K., be regarded as furnishing the proper description of a *perfect liquid* for any purely thermodynamic purpose. (The name *perfect liquid* is due to my colleague, Professor Pitzer.)

In order to appreciate more fully the significance of this device for describing liquid structure, let us note that for a gas W would rise from zero at the distance of closest approach on collision to unity, where it would remain. For a crystal there would be a succession of tall, narrow bands, at the distances at which new groups of molecules are found. The widths of these bands depend on the temperature, melting occurring when they become wide enough to merge into the liquid curve. The liquid curve gradually flattens out with rise of temperature, as shown above for mercury, ultimately reaching and remaining at unity for a gas. The areas under all these curves must, of course, be identical, since the excursions of W above unity must be offset by equivalent dips below.

Let us see what can be got from this method of attack. We can apply it, first of all, to obtain the relation of the potential between pairs of molecules to

the molecular potential of the whole liquid, which, in the case of a perfect liquid, is equivalent to its energy of vaporization. The task is simplified by the fact, theoretically established by London, that these potentials are additive in the case of non-polar molecules. The method differs from that previously applied to the lattice energy of crystals only in that, instead of a summation of the potentials of all the pairs at the distances between all lattice points, with liquids we must perform an integration over numbers of molecules and distances which vary continuously, as given by the experimental form of W.

The number of molecules paired with a certain central one at a distance r in a shell of thickness dr is $4(N/v)Wr^2dr$, and the number of molecules in 1 mole of liquid to serve in succession as central is N, hence the total number of pairs is $(2\pi N^2/v)Wr^2dr$. If the potential between a single pair of molecules is ϵ, also a function of r, the potential energy of the whole liquid per mole is $(2\pi N^2/v)\int \epsilon Wr^2dr$. This should be the energy of vaporization in the absence of intramolecular changes.

The potential may be split into an attractive potential, which London showed to arise from zero-point energy and to vary with an inverse sixth power, and a repulsive potential assumed to vary inversely with some higher power, n. The actual repulsive function is doubtless more complicated, but it will not pay at the moment to attempt to be more precise, so we shall be content to write $\epsilon = j/r^n - k/r^6$.

Professor Lennard-Jones has made good use of this potential function. Substituting this in the previous equation gives

$$E = \frac{2\pi N^2}{v}\left[j\int Wdr/r^{n-2} - k\int Wdr/r^4 \right].$$

If we have the experimental form of W and the heat of vaporization at three or more temperatures, we can, by graphic integration, determine the values of k, j and n. We actually did this for mercury from the x-ray result of my students, Drs. Boyd and Wakeham (Hildebrand *et al.*, 1939e), obtaining $k = 3{\cdot}52 \cdot 10^{-10}$; $j = 5{\cdot}49 \cdot 10^{-9}$ and $n = 9$. The potential curve plotted from these values is shown in figure 6. An independent check is furnished by calculating the attractive constant, k, according to the London formula by aid of the polarizability and the ionization potential of mercury, which yield the value $k = 3{\cdot}35 \times 10^{-10}$, in very gratifying agreement with the above value. I may add that Campbell has recently redetermined the structure of mercury by more precise methods (Hildebrand and Campbell, 1943a), using monochromatic radiation, but a corrected potential curve awaits the end of the war.

A plot of W/r^4 against r for mercury is shown in figure 1C. The area under this curve is approximately proportional to the energy of vaporization. The plot brings out beautifully how little is contributed by molecules beyond the most immediate neighbours.

I shall return later to extend the application of the W function to solutions, but first let us pause to consider molecules which, either in form or force-field, are not spherical. A lack of spherical symmetry in either sense affects the structure and, therefore, the entropy of the liquid. In a perfect liquid the spherical molecules are in a state of maximum disorder consistent with the

volume at their disposal. In a liquid composed of long molecules, such as those of a straight-chain paraffin, there is a tendency for the molecules to assume a parallel array, which introduces a certain degree of order. On evaporation into a large space, both kinds end up with the complete disorder of the gaseous state, dependent only on the volume. The long molecules can thus lose more order on evaporation than the spherical molecules. To test this, let us apply a

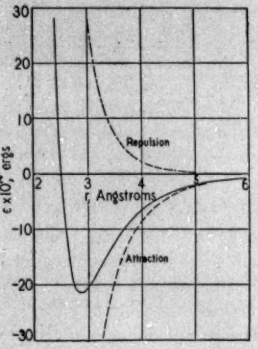
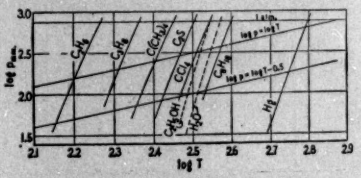


Figure 6.

variation on Trouton's rule which I discovered many years ago. Trouton's rule states that the entropy of vaporization at the ordinary boiling point is the same for a great many liquids, called "normal liquids". The revised version equates entropies of vaporization not at the ordinary boiling points, where the pressures are equal, but at temperatures where the normal volumes of vapour are equal. The distinction is illustrated in figure 7, where $\log p$ is plotted



Uniformity of entropy of vaporization of normal liquids at identical vapour concentration.

Figure 7.

against $\log T$ for liquids covering a wide range of boiling points (Hildebrand, 1939 f). The tangent to a curve at any point, multiplied by R, gives the entropy of vaporization at that temperature, except for deviations from the gas laws, which are small at low pressures. It will be seen that the curves have the same slope, not at equal pressures, but at points cut by a line of unit slope, where the vapours have the same volume. The propriety of expecting equal entropies of

vaporization at temperatures which yield the same volume of vapour, rather than the same vapour pressure, is indicated by the thermodynamic equation $\Delta s = (v_{vap.} - v_{liq.})(dp/dT)$.

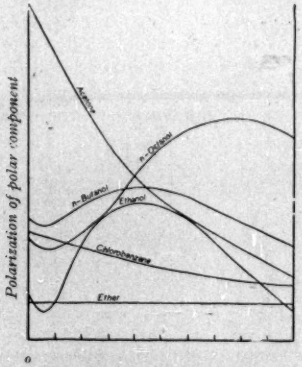
By plotting in this way the rather accurately known vapour pressures of certain liquids, there were obtained the values shown in table 1 for their entropies of vaporization at $\log \boldsymbol{R}/v = 0{\cdot}1$ (Hildebrand, 1939 d).

Table 1

	Δs	$\Delta s - 20{\cdot}1$		Δs	$\Delta s - 20{\cdot}1$
Hg	20·1	0·0	$(C_2H_5)_2O$	21·8	1·7
$C(CH_3)_4$	20·1	0·0	$(CH_3)_2CO$	22·5	2·4
i-C_5H_{12}	20·2	0·1	C_2H_5OH	27·0	6·9
n-C_5H_{12}	20·7	0·6	CH_3OH	26·5	6·4
n-C_6H_{14}	21·5	1·4	CCl_4	20·7	0·6
$(CH_3)_2CH.CH(CH_3)_2$	20·3	0·2	$SnCl_4$	21·8	1·7
$CHCl_3$	21·7	1·6	$C(NO_2)_4$	22·6	2·5

The second column of figures may be taken as indicating the degree of order. Please note the increase in the series neopentane, isopentane, normal pentane; again, with di-isopropyl and normal hexane; carbon tetrachloride, stannic chloride and tetra-nitromethane.

The presence of an electric dipole in a molecule can also lead to a certain degree of order in its liquid, and hence to a higher entropy of vaporization,



Variation of molar polarization with mole fraction.

Figure 8.

and "polar liquids" have long been distinguished from "normal liquids" on the basis not only of this but of other criteria as well. However, the amount of deviation from normal behaviour resulting from dipole moment may be surprisingly small, even for dipoles of considerable magnitude, if the dipole is sufficiently buried. Chloroform and ether have considerable dipole moments,

but the order thus introduced is not enough to add much to the entropy of vaporization. Acetone has a moment not much greater than that of ether, but it is more exposed, and hence introduces more order.

Much larger departures occur in cases where the dipoles are capable of associating through *hydrogen bonds* or *bridges*, as in water, ammonia, hydrogen fluoride, alcohols and carboxylic acids. The extra entropy of vaporization for methyl and ethyl alcohols, given in table 1, far exceeds anything resulting from dipoles alone. This distinction is one of great significance, as we shall see later. Meanwhile, in order to make it more impressive, figure 8 shows the change in molar polarization in going from dilute solutions in non-polar solvents, where the molecules are separated from each other and respond independently to the polarizing field, to ever more concentrated solutions, where they may interact and interfere with each other's response to an external field. Nitrobenzene and acetone behave as we might expect, interacting with each other more and more, and, therefore, less and less with an external field, as they are crowded closer together. Ether, on the other hand, scarcely changes, in harmony with the fact that the dipole is relatively buried under the ethyl groups. The alcohols behave quite differently. Although their polarization drops off at the outset, due to restricted orientation, it then rises enormously. I am convinced that this is to be attributed to the formation of hydrogen bridges wherein, as we now know, the proton can shift between two minima in the potential-well which each occupies between two oxygen atoms. This shift, which accounts for both the entropy and the high dielectric constant of ice, accounts here for the higher polarization as more bridges are formed.

§ 3. SOLUTIONS

Let us turn our attention next to solutions where, by mixing different molecular species, we can enlarge our knowledge of liquids, making further applications of the points of view just set forth.

If we mix two molecular species, having the same size, shape and molecular field strength, the individual molecules are under the same constraints in the solution as in the pure state; hence the tendency of an individual molecule to escape into the vapour phase is unaltered, and the number so escaping is simply proportional to their fraction of the total number present. This is Raoult's law, $p_1 = p_1^0 x_1$, where p_1 is the partial vapour pressure of component one, p_1^0 its vapour pressure in the pure state, and x_1 its mole fraction in the solution. When this relation holds for one component it must hold also for the other. Solutions obeying Raoult's law are called *ideal solutions*.

The change in free energy involved in transferring a mole of one component from its own pure liquid to a large volume of solution is $RT \ln x_1$, from which the change in entropy is $-R \ln x_1$. But let us deduce this in another way, following the well-known procedure for mixed crystals, where, if the two components can participate in the same lattice, the number of arrangements of N_1 and N_2 molecules is $(N_1+N_2)!$ of which $N_1!\,N_2!$ are alike. The probability is $k \ln (N_1+N_2)!/N_1!\,N_2!$, from which, by a few simple operations, we obtain for the entropy of mixing n_1 and n_2 moles (not molecules),

$$R[(n_1+n_2)\ln(n_1+n_2) - n_1 \ln n_1 - n_2 \ln n_2].$$

Now this same analysis is applicable to the entropy of forming a liquid solution, for although the molecules do not occupy definite crystal lattice points, they do possess a definite configuration at any instant in which one species could replace another of the same size and field strength; hence the above expression holds for the entropy of mixing, and from it we can obtain $-R \ln x_1$ for the gain in entropy on transfer of 1 mole from pure liquid to solution, exactly as above. It is interesting to note the magnitude involved. If we form the mixture from 1 mole of each of the components, the increase in entropy is 2·75 calories per degree. Note, also, that in the system we have assumed, with spherical molecules of equal size and field strength, all configurations are equally probable; the disorder is a maximum, and would not be different at a different temperature.

Suppose, next, that we have molecules of different sizes but the same field strength, what happens to Raoult's law? This has not been an easy question to answer, except in the special case of a lattice wherein two spherical molecules could be replaced by a dumb-bell molecule, or some other which is an exact multiple of the first, so that it would occupy an integral number of lattice points. The lattice picture is, however, so artificial that it is gratifying to be able to avoid it for such solutions as those of the normal paraffins, where the attractive forces between adjacent portions of chains are the same, regardless of the lengths of the chains.

Such solutions present a nice puzzle. A person educated in the osmotic school, where the method of expressing concentration is moles per litre, would probably conclude that ethane at one atmosphere would dissolve to about the same extent in a given volume of any higher paraffin, since an ethane molecule is subject only to the attraction of that part of the large molecule which is next to it, and is not affected by any extra length the latter may have (Hildebrand, 1939 c). Raoult's law, however, leads to a different conclusion, for the mole fraction of a given amount of ethane in a certain volume of hexane would be only half what it would be if dissolved in an equal volume of dodecane; hence its pressure would have to be doubled over the latter to give it the same solubility.

Experiment decides unequivocally in favour of mole fraction instead of moles per litre. Raoult's law is closely obeyed by normal paraffins having such different lengths as butane and heptane, and again by hexane and hexadecane (Hildebrand and Sweeney, 1939 c). Indeed, the Gibbs-Duhem equation shows that Henry's law expressed in moles per litre can be true for both components of a solution only when their molal volumes are equal. What is wrong with the explanation based on volume concentration?

Put very simply, it is the spaces *between* the molecules of the solvent that are available to a solute, not the total space. Liquids composed of large molecules are much poorer solvents, other things being equal, than liquids with small molecules, volume for volume. But let us go more fully into one particularly interesting case.

The free energy of mixing can be split into its energy and entropy terms. The energy of mixing normal paraffins is practically zero. The entropy of mixing can be calculated (Hildebrand, 1939 c) if we assume that the paraffin chains are arranged roughly parallel, which must be true at least not too far above the melting point, from the number of arrangements possible for the

mixture of the long and short molecules, when arranged in a long line. It comes out the same, numerically, as the number of arrangements of two species of spherical molecules in a point lattice; hence the entropy of mixing is the same as for spherical molecules, and here again Raoult's law holds, as careful experiments have shown in the case of hexane and hexadecane.

Let us next consider the effect of different field strengths. This introduces a tendency towards segregation, which may go so far as to cause separation into two only partially miscible liquid phases if the temperature is lowered below a point where thermal agitation is insufficient to prevent it. Now one might suppose that even above this point temporary clusters of the species with stronger field should be able to form: and this is the case, the clusters being large enough to be visible in a small region around the critical point. Fortunately, however, this clustering rapidly disappears as we move away from this point (Hildebrand and Negishi, 1937 b), either in composition or in temperature, and we soon reach conditions where the molecular distribution is as random as it is in an ideal solution of the same composition. In such a case, the transfer of a molecule from an ideal solution to the non-ideal solution of the same mole fraction should involve no change in entropy, and the free energy of transfer should be equal to the difference in energy. The entropy is ideal, although the free energy is not. Such solutions show a significant regularity of behaviour which has invited the designation *regular solutions*.

We may now proceed to derive laws for the regular solutions. Let us mix N_1 and N_2 molecules of two perfect liquids having spherical molecules of the same size, but unequal molecular fields, so that three intermolecular potentials are involved, which we shall designate ϵ_{11}, ϵ_{22} and ϵ_{12} respectively, and let us choose a temperature high enough so that molecular agitation is enough to overcome any tendency to segregation due to the unequal attractive forces. The distribution function, W, is the same as for a pure liquid, and the total number of pairs at any distance apart is as given earlier; however, the chances that these pairs are of the same species, 1–1 or 2–2, or of different species, 1–2, are, respectively, $N_1^2/(N_1+N_2)^2$; $N_2^2/(N_1+N_2)^2$; $2N_1N_2/(N_1+N_2)^2$. The potentials of all the pairs of each kind throughout the solution can be expressed in these terms, and, added together, give the potential energy of the solution. The mathematical steps have been published, and I shall not include them here, but merely write the result, aided by neglecting the repulsive potentials and by assuming, as is usual, that the attraction constant for the unlike molecules is the geometric mean between the constants for the like molecules. The energy of transfer of 1 mole of component 1 from pure liquid to solution of mole fraction x_1 is $[(\Delta E_1)^{\frac{1}{2}}-(\Delta E_2)^{\frac{1}{2}}]^2(1-x_1)^2$.

The entropy of transfer is $-R\ln x_1$, and from these we can calculate the free energy of transfer. Iodine and carbon disulphide have nearly equal molal volumes in the liquid state, 59·0 c.c. (extrapolated) and 60·7 c.c., respectively. Their molal energies of vaporization at 25° are, respectively, 11,300 cal. and 5970 cal. From these values the quantity in brackets above is 840 cal. From the solubility of iodine in carbon disulphide we obtain 870 cal., a very gratifying check, in spite of the fact that their molecules are not spherical.

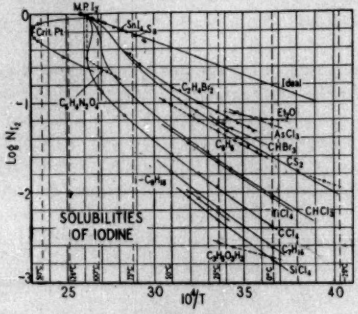
If the molecules are of different sizes, the matter is rather more complicated,

and to arrive at a practical equation involves a short cut or two. The method has been published and is available for scrutiny and criticism, so I here jump straight to the conclusion and to the check against experiments:

$$RT \ln \frac{a_1}{x_1} = v_1 \left(\frac{x_2 v_2}{x_1 v_1 + x_2 v_2} \right)^2 \left[\left(\frac{\Delta E_1}{v_1} \right)^{\frac{1}{2}} - \left(\frac{\Delta E_2}{v_2} \right)^{\frac{1}{2}} \right]^2 . \qquad \ldots\ldots(1)$$

The symbol a denotes the activity and equals the mole fraction the substance would have in an ideal solution, which is readily calculated from its heat of fusion. Remember that it is not supposed to apply to molecules with queer shapes or unsymmetrical fields, or where selective or chemical intermolecular forces are operating. If it does fit such cases, as, indeed, it often does, that is just good luck.

Figure 9 shows the solubility of iodine (Hildebrand, 1939 f) in a variety of solvents plotted, appropriately, as the logarithm of the mole fraction of iodine against the reciprocal of the absolute temperature. The line at the top represents



Solubilities of iodine.

Figure 9.

the solubility in an ideal solution, where Raoult's law holds, and the heat of solution is identical with the heat of fusion of iodine.

We see that most of the curves belong to a family justifying the designation *regular*, given above, and corresponding to ideal entropy but different energies and free energies of solution. All of these solutions have practically the same colour, which is that of iodine vapour, indicating that the iodine molecules are not chemically altered. Although the deviations from the ideal line are mostly very large, it is obviously ridiculous to attribute them to the formation of new molecular species.

But some of the curves clearly do not belong to the regular family. It is significant that all of these solutions have colours quite different from the violet iodine vapour, reds, yellows and browns, indicating that the I_2 molecules have suffered changes which we may properly call chemical.

The colour difference between the regular solutions and the others, in the case of iodine, makes it easy to see that large deviations from ideal solubility or, what amounts to the same thing, lowering of freezing point, may be either

"physical" or "chemical", due either to difference in molecular field-strength on the one hand or to the formation of new molecular species on the other. The latter is too often assumed when the former is actually responsible.

The equation accounts very well for the family of "violet" curves, despite the approximations that had to be made in its derivation. The correspondence between observed and calculated solubilities is illustrated by the following:—

Table 2. Solubility of iodine (25°)

Solvent	Mole per cent iodine	
	Calculated	Observed
Ideal	21·2	—
CS_2	5·5	5·8
$TiCl_4$	1·7	2·1
CCl_4	1·6	1·4
$SiCl_4$	0·4	0·5

Another way of making the test is to substitute in the equation the observed solubility, the molal volumes, and the value of $\Delta E_1/v_1$ for the solvent and calculate the value of $\Delta E_2/v_2$ for iodine, comparing with the value derived directly from the vapour pressure of liquid iodine. We obtain values such as the following:—

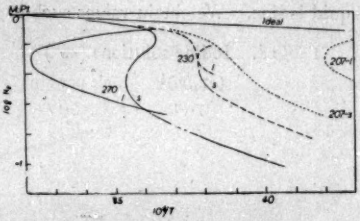
Table 3. Iodine solutions (25°)

Solvent	$(\Delta E_1/v_1)^{\frac{1}{2}}$	Mole per cent I_2	$(\Delta E_2/v_2)^{\frac{1}{2}}$
n-Hexane	7·44	0·47	13·67
Silicon tetrachloride	7·57	0·50	13·71
Carbon tetrachloride	8·54	1·15	13·99
Chloroform	8·93	2·28	13·71
Carbon disulphide	9·98	5·76	13·82
Bromoform	10·10	6·16	13·78
Ideal	13·60	21·2	13·60

Returning to the plot, figure 9, you will see a loop in the upper left-hand corner intersecting the curve for carbon tetrachloride. It is perhaps not a very important brain-child, but I had an unusual amount of fun in being its parent. The equation shows that the curve running through the points for the solubility of solid iodine must take on the reversed S-shape shown higher up. The significance of this is like that of the similar curve obtained on plotting the van der Waals equation, that is, there are two compositions in equilibrium at a single temperature. In this case it is due to the incomplete mutual solubility of liquid iodine and carbon tetrachloride. The solid-liquid and the liquid-liquid curves are related thermodynamically, so that, knowing one, we can calculate the other. I will not burden you with details of the calculation, published elsewhere, but simply say that a consolute temperature was predicted somewhere between 150° and 170° at 71 mole per cent iodine.

But iodine solutions so concentrated are quite opaque, even in fine capillaries, so that the existence of two liquid phases between the melting and the consolute temperatures could not be verified by visual observation: hence an indirect method had to be used. The two substances in about the right proportions were sealed up in a glass tube; this was balanced at a small angle to the horizontal in a cradle provided with a small mirror and held by a spring, the whole in a

hot-air bath. A beam of light was reflected from the mirror on to a scale. The bath was heated to about 180°, sufficient to give one liquid phase, and then allowed to cool slowly, while readings were taken of the light pointer on the scale. As soon as the liquid broke into two phases, the iodine-rich one, being much denser, ran to the bottom of the tube, exerting a moment which caused the light spot to shift rapidly as the temperature was lowered further. The break in the scale-temperature curve, when plotted, was very sharp. A series of tubes with fillings of different compositions gave the complete liquid-liquid solubility curve, with a maximum at 160°·5, within the predicted range, and a composition at that point of 68 mole per cent, agreeing well with the calculated value of 71 per cent. The fact that the whole course of such a liquid-liquid solubility curve can be calculated with but little error from the equation for regular solutions illustrates the minor rôle of clustering near the critical point. It seems fair to assume that the existence of this particular liquid pair would never have been suspected without the aid of some such theoretical approach. A similar behaviour has been found in solutions of stannic iodide with silicon tetrachloride.



Relation between solubility curves for solid and liquid forms.

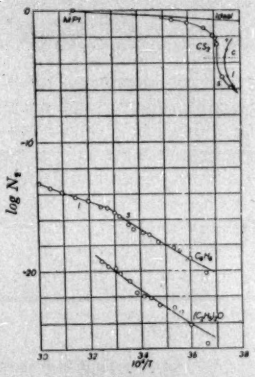
Figure 10.

A striking case is furnished by the liquid-liquid system, stannic iodide and dicetyl, $C_{32}H_{66}$, which has a consolute temperature of 194°. The composition at this point is 90 mole per cent of stannic iodide, far removed from the usual region of about 50 per cent. Now the shape of the long dicetyl molecule is far from the spherical assumed in deriving our equations, unless, indeed, the long chain is curled up in this case, and one might reasonably expect the calculated composition to be pretty far from the measured. The agreement is, however, surprisingly good. The molal volumes of the two liquids at the consolute temperature are 171 c.c. and 615 c.c., respectively, from which we calculate for the consolute composition 85 mole per cent of stannic iodide, only 5 per cent away from the measured value.

It is desirable to put the maximum strain on the theory by choosing substances giving as large deviations as possible from ideal behaviour. Active phosphorus, whose molecules are P_4, and are very compact tetrahedra, has a field strength so high that it forms two liquid phases with most of the familiar organic solvents, even those which are non-polar. Indeed, carbon disulphide, one of its best solvents, gives two liquid phases with supercooled liquid phosphorus.

Figure 10 shows the types of solid and liquid solubility curves for phosphorus

to be expected for three different solvents of different molecular fields or internal pressures. The curves at the left are similar to those for iodine and carbon tetrachloride, previously shown, where two liquid phases coexist above the melting point of phosphorus. The curves in the middle correspond to phosphorus and carbon disulphide, where the solid and liquid curves are tangent at the critical point of the latter. The curves at the right represent a more nearly ideal solution, where great super-cooling would be necessary to reach two liquid phases. Figure 11 shows the solubility curves for phosphorus in ether and benzene, which illustrate the pair at the left of the preceding figure, where solid and liquid curves intersect, together with the solid and liquid curves for carbon disulphide. When I first plotted the points for solid phosphorus, I had not yet plotted the theoretical curve, and was puzzled by their course, unlike anything I had previously encountered. It was a pleasant surprise to find that they fitted the equation so closely, in harmony with the near-by liquid-liquid



Solubilities of phosphorus.

Figure 11.

curve that one of my students had determined years before. All of these curves for solid and liquid with the three solvents agree quantitatively with the theory within reasonable limits of error.

Dipole moment often makes less difference than one might expect. Thus, iodine gives a regular and violet solution with chloroform, in spite of its large dipole moment. A more striking case is presented by the solutions of nitro-substituted benzenes (Hildebrand, 1939f), represented in figure 12, where excess free energy of mixing over that for an ideal solution for each of the several nitrobenzenes is plotted against the square of the mole fraction of benzene, a method which eliminates the differences due to different heats of fusion. The appendages indicate position and number of the nitro groups. It will be seen that all of the di-nitro benzenes, ortho, meta, and para, fall on the same line, despite their widely different dipole moments. Comparison of the lines for mono-, di- and tri-nitro benzenes show that the deviation from Raoult's law

depends only on the number, not the position, of the nitro groups. The dipole vector has little or nothing to do with it.

A comparison of anthracene and phenanthrene is interesting, for these two isomers differ only in shape, the former consisting of three hexagons joined by edges in a line, the latter bent. The result is that anthracene packs much better into crystals, with much higher density and melting point, as shown in table 4. It is, therefore, more difficult to detach molecules from the crystals to dissolve in benzene. This difference is, however, due almost solely to the difference in the two solids, for, in the liquid state, thermal agitation is enough to make their molecules exert the same average external field, as shown by their identical boiling points.

Table 4

	Molal vol. (c.c.)	Melting point (°c.)	Boiling point (°c.)	Solubility in benzene (mole per cent at 25° c.)
Phenanthrene	174	99·6	340	18·6
Anthracene	142	218	342	0·63

Let us return now to a brief consideration of solutions of substances capable of forming hydrogen bridges, which add very considerably to the forces of attraction between the molecules thus joined. These bonds are not easily broken by molecules of another species unless the latter can participate in this special structure; accordingly, water, hydrogen fluoride, ammonia and the alcohols are much poorer solvents for non-polar and non-ionic solutes than they would otherwise be. They truly merit the term *associated*. Their "polarities" or dipole moments alone do not serve to account for their behaviours. This is strikingly illustrated by the figures in the following table:—

Table 5. Solubility of liquids in water at 20°

	Dipole moment	Mole per cent.
Ethyl iodide	1·66	0·40
Ethyl alcohol	1·70	∞
Benzene	0	0·06
Nitrobenzene	4·19	0·19
Aniline	1·51	3·49
Phenol	1·70	8·20

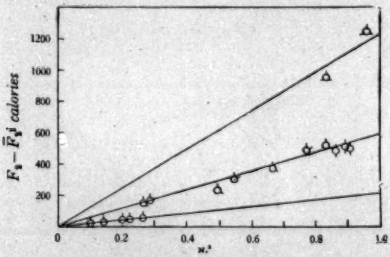
Ethyl iodide and ethyl alcohol have nearly identical dipole moments, but the former is only slightly soluble in water, while the latter mixes with water in all proportions. Again, the increasing solubilities of benzene, nitrobenzene, aniline and phenol correspond well with their relative abilities to form hydrogen bridges, but not at all well with their dipole moments. Again, the solubility of naphthalene is not far from ideal in nitrobenzene, despite the large dipole moment of the latter, less in acetone but very much less in alcohol, although the latter is less polar. Naphthalene molecules cannot easily break the hydrogen bridges binding together the molecules of alcohol.

Table 6. Solubility of naphthalene at 20°

	Dipole moment	Mole per cent.
Ideal	0	26·1
Chlorobenzene	1·56	25·6
Nitrobenzene	4·19	24·3
Acetone	2·80	18·3
Butyl alcohol	1·68	4·95
Methyl alcohol	1·68	1·80

The evident necessity of taking hydrogen bridging into account in these relationships is in striking contrast to its neglect in the current theories of dilute electrolytic solutions, where solvent water is treated simply as so much space with a dielectric constant. Perhaps ionic theory could be pushed farther into the region of concentrated solutions if more regard were paid to the actual structure of liquid water.

A considerable amount of accurate data regarding metallic solutions has been obtained by measuring the e.m.f. of alloy concentration cells, free from



Relation between number of nitro-groups in nitrobenzenes and deviation from Raoult's law in benzene solutions.

Figure 12.

liquid junctions, over a wide range of temperature, permitting the calculation of partial molal free energies, entropies and heats. Some of them turn out to be "regular" in the sense of having the same entropy as ideal solutions of the same molal composition, as is the case with so many non-metallic solutions. They differ from the latter, however, in not exhibiting additive inter-molecular potentials, so that one cannot calculate the free energy from values of energy of vaporization per c.c. of the pure components. We can test this without knowing these individual values, which are not easily measured, by considering a triangle of components and calculating the values that would have to be assigned to the factor $(\Delta E_1/v_1)^{\frac{1}{2}} - (\Delta E_2/v_2)^{\frac{1}{2}}$ in equation (1) in order to give the experimental free energies. They are as follows for the binary solutions of Cd, Sn and Zn: Cd–Sn ± 8·22; Cd–Zn ± 13·05; Sn–Zn ± 12·04. It is evident that no pair can be got either by adding or subtracting the other two. It is quite different for non-metallic solutions. Solutions of CS_2, C_6H_6 and SnI_4 give the following:

CS_2–C_6H_6, 1·82; CS_2–SnI_4, 0·96; C_6H_6–SnI_4, 2·92; the last is close to the sum of the first two, 2·78.

The non-additivity in the case of metallic solutions is hardly surprising in view of the fact that zero-point energies of atoms, which are additive, do not provide all the forces operating, the free electrons adding their own large effects. I can only say that accurate thermodynamic data are on hand awaiting theoretical treatment.

§ 3. CONCLUSION

In closing, let me give an illustration indicating at once both the progress that has been made and the long way still to go. Figure 13 depicts a system consisting of six liquid layers, stable with respect to each other: mercury, gallium, phosphorus, water, aniline, hexane. The fact that so many liquids are immiscible, or only partially miscible, in contact with each other illustrates the variety possible in the kinds and the strengths of molecular fields. Now, the methods outlined in this lecture permit a rather

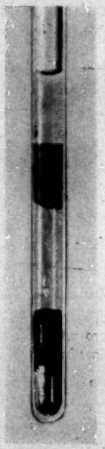


Figure 13. Six immiscible liquids.

good quantitative calculation of the mutual solubilities of phosphorus and hexane, and they explain why mercury and gallium are immiscible in the other liquids. They suffice to explain only qualitatively the relative solubilities of hexane and aniline in each other and in water. They yield no adequate explanation of the mutual insolubility of gallium and mercury. I trust that this illustration conveys both a fair impression of the progress that has been made and some notion of the unknown regions inviting exploration whenever we are again free to turn our attention away from munitions of war.

REFERENCES

EISENSTEIN, A. and GINGRICH, N. S., 1942. "The diffraction of x rays by liquid and gaseous argon". *Phys. Rev.* **62**, 261.

GINGRICH, N. S., 1943. "The diffraction of x rays by liquid elements". *Rev. Mod. Phys.* **15**, 90.

HILDEBRAND, J. H., 1936 a. "Solubility of non-electrolytes", *Amer. Chem. Soc. Monograph*, 2nd edition (New York, Reinhold Pub. Corp.; (with W. E. MORRELL) 1936 b, "The distribution of molecules in a model liquid", *J. Chem. Phys.* **4**, 224; 1937 a, "Intermolecular forces in solutions", *Trans. Faraday Soc.* **33**, 144; (with G. R. NEGISHI) 1937 b, "The solubility of liquid and solid stannic iodide in silicon tetrachloride", *J. Amer. Chem. Soc.* **59**, 339; 1937 c, "The validity of Raoult's law for molecules of very different lengths", *ibid.* **59**, 794; 1937 d, "The incomplete solubility of liquid iodine in carbon tetrachloride", *ibid.* **59**, 2083; 1939 a, "Several solutions of non-polar substances", *J. Phys. Chem.* **43**, 109; (with J. W. SWEENEY) 1939 b, "The entropy of solution of hexane with hexadecane", *ibid.* **43**, 297; 1939 c, "Liquid structure and energy of vaporization", *J. Chem. Phys.* **7**, 1; 1939 d, "Liquid structure and entropy of vaporization", *ibid.* **7**, 233; (with H. R. R. WAKEHAM and R. N. BOYD) 1939 e, "The intermolecular potential of mercury", *ibid.* **7**, 1094; 1939 f, "Order and disorder in liquids" (Nichols Medal Address), *Science*, **90**, 1; (with J. A. CAMPBELL) 1943 a, "The structure of liquid mercury", *J. Chem. Phys.* **11**, 330; (with J. A. CAMPBELL), 1943 b, "The structure of liquid xenon", *ibid.* **11**, 334.

KIRKWOOD, J. G. and BOGGS, E. M., 1942. "The radial distribution function in liquids". *J. Chem. Phys.* **10**, 394.

THE MOBILITIES OF ELECTRONS IN HELIUM

BY F. LLEWELLYN JONES, M.A., D.PHIL., F.INST.P.,
University of Bristol

MS. received 4 *April* 1944

ABSTRACT. The electron mobilities in helium are calculated over the range of Z/ρ from 0·5 to 3·0 v./cm./mm. Hg on the basis of an electron energy distribution formula which has previously been found to be applicable over the electron energy range required for excitation and ionization in the uniform positive column in helium. When the observed variation of atomic cross-section with electron velocity is taken into account, the calculated mobilities show good agreement with the accepted measured values. It is suggested that this agreement may be regarded as evidence in support of the distribution formula considered.

§1. INTRODUCTION

A CALCULATION of the mobility of electrons in gases is of particular interest because it involves the relation of elementary atomic collisional processes to macroscopic quantities which can readily be observed. The measurement of electron drift currents in gases under uniform electric fields is of importance, in the first place, because a knowledge of them is necessary in many investigations of electrical discharges through gases, and, in the second place, because a knowledge of them affords a test of the validity of the fundamental assumptions relating to the elementary processes (see, for instance, Loeb 1939, and Francis and Jenkins, 1940).

The function which relates the elementary processes to the final expression for the mobility of an electron group is the electron energy distribution function; this is one of the reasons for the great importance of that function. In this paper the interest in the exact form of the expression for the electron mobility

lies in its variation with the form of the particular energy distribution assumed, and also on the effect of the variation of the atomic cross-section with the velocity of the colliding electron.

It has been shown in a previous paper (Llewellyn Jones, 1936) that the optical properties of discharges in helium in a uniform electric field can be accounted for in terms of the electron energy distribution if the distribution of the electron energies E about the mean energy E_1 is very nearly given by the equation

$$dn = N(Ay^{1/5}e^{-0\cdot4y^2} + By^{1/2}e^{-3y/2})\,dy, \qquad \ldots\ldots(1)$$

where

$$A = 0\cdot7736, \quad B = 3\cdot896 \times 10^{-3}, \quad y = mu^2/mu_1^2 = E/E_1.$$

This expression differs a little from that originally given, which was meant to apply only to the distribution of the high-energy groups of electrons responsible for excitation and ionization. The expression in equation (1) can apply over the entire energy range, and is the complete distribution formula. It will be discussed in more detail in another paper.

In the processes of excitation and ionization the only relevant part of the distribution curve is that in which the electron energies exceed the lowest excitation level; and the number of electrons in that region is only a very small fraction of the total number. Consequently the distribution function of those electrons which have energies not very different from the mean energy, up to about three times the mean energy for example, is not of great importance when considering the processes of excitation and ionization and their dependence on the gas density in discharges at pressures exceeding about 1 mm. Hg. On the other hand, when considering the mean drift velocity of the electrons, the distribution of the energies of those electrons which have energies lower than the resonance energy is of importance, because these electrons constitute a very large fraction of the total number. Hence it is possible to test the validity of a distribution formula such as that given in equation (1), more especially over the lower energy ranges, by using it to calculate the electron mobility in helium for different values of Z/ρ, the ratio of the electric force Z to the gas density ρ in mm. of Hg, and then to compare those values with the experimental determinations.

The distribution formula in (1) is similar to that which has been found to agree with experiments on excitation and ionization in helium (Llewellyn Jones, 1936), and it is therefore considered to represent the distribution which actually occurs in helium when the ratio Z/ρ is less than 6·0 v./cm./mm. Hg. This range of Z/ρ corresponds to the régime under which most mobility measurements have been made, and the positive column most studied. Since it is assumed that the electron velocities are given by equation (1), it is then of interest to calculate the mobility by free-path methods, introducing the variation of atomic cross-section $Q(E)$ with the electron energy E, and thus to express the electron mobility as far as possible in terms of the directly observable parameter Z/ρ and the fundamental atomic constants only. The comparison of the calculated values with the experimental determinations of the electron mobilities consequently serves as a test of the validity of the distribution formula, especially over that part of the energy range which is unimportant in excitation and ionization, and which, therefore, cannot be easily investigated by spectroscopic or ionization

methods. The calculation applies to the régime of elastic collisions between the electrons and atoms, or to those conditions when the losses due to excitation and ionization are small compared with the elastic losses. This régime of elastic conditions, which in helium approximately corresponds to the cases when Z/p is less than 6, includes a large range of gas pressures at which positive columns can be observed (Llewellyn Jones, 1931).

However, before proceeding with the expression of the electron mobility in terms of atomic constants only, it is of interest to consider the formula for the mobility in terms of the root-mean-square electron velocity in the classical form first given by Langevin (1905). This particular formula is often used in applications to electrical discharges, and it is also useful when considering how the mobility depends on the form of the electron energy distribution. The formula given by Langevin for the case of charged particles of mass m, which is small compared with the mass M of the atoms of the gas through which they move under a very weak electric field of intensity Z, is (Loeb, 1939)

$$W = heZL/mpu_1,$$

where

L = mean free path of the particle in the gas at a density corresponding to a pressure of 1 mm. Hg at 15° c.,

e = charge on the particle,

u_1 = root-mean-square velocity of the particle,

h = numerical coefficient.

This constant h depends on the energy distribution of the particles, and Langevin showed that when the distribution was practically Maxwellian, and not very different from that of the gas atoms, then the value of h was 0·815.

Since the electron energy distribution varies with the type of electrical discharge, it is of considerable interest first to investigate how h varies with the form of the energy distribution, at the same time assuming a constant ideal gas-kinetic cross-section. Later in the paper the value of the coefficient h is calculated for a certain energy distribution (equation (1)) when the effect of the variation of the atomic cross-section is taken into account.

§ 2. THE ELECTRON MOBILITY AND VARIOUS ENERGY DISTRIBUTIONS

The method here adopted for calculating the mobility, first with constant atomic cross-section, is the free-path method of Townsend (1936). The steps in the derivation are given because the same method will be used later when considering the effect of the variable cross-section on the mobility.

Consider a group dn of electrons which have velocities within the range u, $u+du$, and which are scattered uniformly by a point source—a gas atom. Those electrons with velocity vectors making angles lying between θ and $\theta+d\theta$ with the direction of the electric field Z, supposed weak, are displaced along the field during a free path a distance

$$\gamma \sin^2\theta \,.\, t^2/2,$$

where $\gamma = eZ/m$ and t is the interval between successive collisions with atoms when the electron speed is u. The scattering of electrons by positive ions may

be neglected in comparison with that due to atoms (Druyvesteyn and Penning, 1940). Consider the case when the electron energies are large compared with those of the gas atoms; in helium, for example, this occurs when Z/ρ exceeds about 1·5 (Llewellyn Jones, 1931). Further, suppose that all directions of scattering are equally probable, as is the case in an ideal gas.

Now the free paths of the electrons in this particular (u, θ) group are distributed about the mean free path, so that the average drift speed $W(u, \theta)$ of this group along the electric field is given by the distance traversed per second along the field when averaged over a large number of free paths. Hence

$$W(u,\theta) = \gamma \sin^2\theta \,.\, \sum_0^\infty t^2/2 \sum_0^\infty t = \gamma \sin^2\theta/\rho u Q(u),$$

where $Q(u)$ is the total atomic cross-section for ideal elastic scattering at a gas density corresponding to a pressure of 1 mm. Hg at 15° c.

To find the average drift speed $W(u)$ of all the electrons with velocities within the range $u, u+du$, this expression for $W(u,\theta)$ must be averaged over all values of θ. Thus (Townsend, 1936)

$$W(u) = 2eZ/3m\rho u Q(u).$$

This is the drift speed of the electron group averaged over a number of free paths during which the speeds of the electrons are altered only by an infinitesimal amount by the electric field, so that they still remain substantially within the range $u, u+du$.

The mean drift velocity $\overline{W}$ of all the electron velocity groups which are distributed about the root-mean-square velocity u_1 is obtained by taking the average of $W(u)$ over all possible velocity groups. Hence

$$\overline{W} = (2eZ/3m\rho) \,.\, \int_0^\infty \{u \,.\, Q(u)\}^{-1} \,.\, dn \Big/ \int_0^\infty dn. \qquad \ldots\ldots(2)$$

But $Q(u)$ is taken to be a constant independent of u, and equal to $1/L$, where L is the electron mean free path in the gas at a density which corresponds to a pressure of 1 mm. at 15° c., so that

$$\overline{W} = (2eZL/3m\rho) \,.\, (\overline{1/u}).$$

Now it is convenient for many applications of this formula to express it in terms of the root-mean-square velocity u_1. In that case, the numerical coefficient is no longer 2/3, but has some other value h, which depends on the form of the electron energy distribution. Thus

$$\overline{W} = h \,.\, eZL/m\rho u_1. \qquad \ldots\ldots(2')$$

This expression for the electron mobility corresponds more closely to what is measured experimentally than does the instantaneous speed of the electron group averaged over all the electrons.

To calculate the coefficient h it is necessary to know dn in terms of u.

Consider the following energy distributions which are of interest for their application in many types of discharges. These, in order of flatness, are:

(*a*) An infinitely narrow distribution with all velocities equal to u_1. This could only occur in practice for a very short time in an ideal gas, but the influence

of the Townsend-Ramsauer effect in some gases, notably argon, can produce very narrow distributions (McCallum and Klatzow, 1933). In such cases it is often sufficiently accurate to assume that all the electron velocities are equal.

(*b*) The diffusion distribution of Druyvesteyn (1930):

$$dn = C \,.\, N \,.\, y^{1/2} e^{-0{\cdot}55y^2} \,.\, dy.$$

This occurs in an ideal gas when the mean electron energy is of the order of 100 times that of the gas atoms, and only elastic energy losses are important.

(*c*) Maxwell's distribution. This can occur in light gases under very weak electric fields, and possibly in polyatomic gases in which considerable energy interaction can occur between the molecules and colliding electrons owing to the occurrence of low-energy-value vibrational levels. It may also occur with very high electron densities under conditions when substantial energy interchanges take place between the electrons themselves.

(*d*) The distribution given in equation (1) deduced from the observed optical and electrical properties of helium. This is broader than (*b*).

The values of h, calculated by means of equations (2) and (2′), corresponding to these distributions, are given in table 1, from which it can be seen that the

Table 1

Distribution	Mobility coefficient h	Coefficient k
a. All u's equal	0·66	3
b. Druyvesteyn	0·83	2·73
c. Maxwell	0·92	2·66
d. Equation (1)	0·907	2·64
e. Equation (1), but considering variable $Q(E)$ for helium, $E_1 = 4{\cdot}0$ ev.	0·893	2·23

mobility coefficient diminishes as the distribution becomes narrower. The values in the last row are those which correspond to $Z/\rho = 2$. They are different for other values of Z/ρ.

The last column of the table gives values of a certain constant k which is discussed below.

The value of the numerical coefficient h corresponding to the Druyvesteyn distribution is the same as that obtained by Morse, Allis and Lamar (1935).

The form of the equation (2) is useful when the r.m.s. value u_1 of the electron velocity is known. In general, this is not the case, so that it is interesting to eliminate u_1 by expressing it in terms of the observable parameter Z/ρ and the atomic constants. This may be done as follows.

The energy equation for a stream of electrons of concentration N moving under a uniform electric field Z in an ideal gas at density ρ when ionization and excitation are neglected and elastic losses predominate is

$$d(NE_1)/dt = \overline{W} \,.\, NeZ - \int_0^\infty \lambda E^{3/2} Q(E) \,.\, \rho \,.\, dn/(m/2)^{1/2}, \qquad \ldots\ldots(3)$$

where λE is the mean loss of energy of the electrons of energy E in elastic collisions with gas atoms. This equation is a statement of the fact that the rate of increase of the mean energy of the electrons is equal to the difference between the rate of gain of energy from the electric field and the rate of loss of energy due to the losses in elastic collisions.

At equilibrium

$$d(NE)_1/dt = 0,$$

so that

$$2\overline{W}eZ/\lambda = \int E^{3/2}Q(E) \,.\, \rho \,.\, dn/N(m/2)^{1/2}. \qquad \ldots\ldots(4)$$

Using the expression for $\overline{W}$ given in equation (2′), and remembering that Q is taken to be constant, it can be seen that

$$4e^2Z^2h/2\lambda Q^2(m^2u_1^4) = \int (u/u_1)^3 \,.\, dn/N = c,$$

where c is a constant (≈ 1) whose value depends on the form of the distribution function. The mean electron energy E_1 can thus be expressed as

$$eE_1 = mu_1^2/2 = eZ/\rho Q\sqrt{(k\lambda)}, \qquad \ldots\ldots(5)$$

where k is a constant ($= 2c/h$) whose value depends on the distribution.

Now the coefficient λ can be expressed in terms of the ratio m/M of the electronic and atomic masses, and the form of the distribution function (Druyvesteyn and Penning); thus

$$\lambda = gm/M, \qquad \ldots\ldots(6)$$

where g is a constant whose value depends on the distribution. For instance, g is 2 for an infinitely narrow distribution, and is 2·66 for Maxwell's distribution (Cravath, 1930), so that it is probable that g does not differ very much from 2·66 for the other distributions given in table 1, since these do not differ much from the Maxwellian form.

On the basis of these assumptions the value of the coefficient k has been evaluated for the four types of distribution considered above, and the values are given for comparison in table 1 (a to d). These values of k, together with those of h, serve to illustrate how the mobilities and the mean energies of agitation of the electrons depend on the distribution law when the atomic cross-section is constant.

However, in order to calculate the electron mobility in helium, and so to test the validity of the distribution formula given in equation (1) over the complete energy range, including the lowest energies, it is necessary to use equations (1), (2) and (4), at the same time taking into account the observed variation of atomic cross-section with velocity: the calculated values can then be compared with those observed.

§ 3. THE ELECTRON MOBILITY IN HELIUM

In applying equation (2) to the case of helium, it is necessary to note that the left-hand side involves the atomic cross-section for elastic scattering, while the right-hand side involves the cross-section for momentum transfer. As Gvosdover has pointed out (Francis and Jenkins, 1940), these two cross-sections may not be the same, and the difference between them depends upon the degree

of any preferential scattering in any given direction. Again, when the scattering is non-uniform, the mean loss of energy of an electron in an elastic collision with a gas atom involves the term

$$\int_0^\pi f(\theta)(1-\cos\theta)\,d\theta,$$

where $f(\theta)$ is the probability that the direction of the electron after collision makes an angle θ with the original direction (Druyvesteyn and Penning, 1940). In consequence the mean electron energy is no longer given by equation (5), but it can be shown that

$$eE_1 = eZ/\rho Q_0 \sqrt{k\lambda(1-P)},$$

where

$$P = \overline{\cos\theta}.$$

Thus P is zero for an ideal gas, since the scattering is uniform, and as the non-uniformity of the scattering in helium is not great, P is practically zero in helium also. Hence, it may be concluded without serious error that equation (5) may

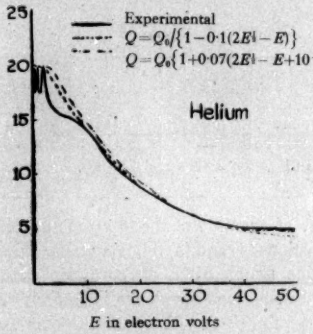


Figure 1.

be applied to the case of helium. However, although the atomic cross-section may thus be taken to be independent of the angle of scattering, it still depends on the velocity of the colliding electron.

Experimental investigations of the atomic cross-section have been made by Normand (1930) and by Ramsauer (1929), but in order to evaluate the integrals required in equations (2) and (4),

$$\int \{u\,.\,Q(u)\}^{-1}\,.\,dn \quad \text{and} \quad \int E^{3/2}\,.\,Q(E)\,.\,dn,$$

and avoid tedious expressions, it is convenient to express $Q(E)$ or $Q(u)$ as far as possible as a simple function of E. Suitable functions are:

$$Q(E) = Q_0\{1-0{\cdot}1(2E^{1/2}-E)\}^{-1}, \qquad \ldots\ldots(7)$$

$$Q(E) = Q_0\{1+0{\cdot}07(2E^{1/2}-E+10^{-2}E^2)\}. \qquad \ldots\ldots(8)$$

These two functions are graphed for comparison in figure 1, together with a curve which approximates to the experimental results of Ramsauer and Normand;

Q_0 being taken as 18·8. It can be seen that the agreement is sufficiently good for the mobility calculations. The departure of equation (8) from the experimental values at high values of E is of little consequence owing to the rapid decrease of dn at those extreme values of E. In order to find whether the mobility coefficient depended to any great extent on the exact detailed form of the variation of $Q(E)$ with E, a number of slightly different expressions for $Q(E)$ were used in the calculations. In that way it was found that the numerical values of the coefficient h did not depend very much on the detailed variation of $Q(E)$ with E, but only on the general trend of the function. From this it was considered that any errors due to the use of the simple functions given in equations (7) and (8) (instead of the use of the exact function, if this were in fact known correctly) could be neglected in the integrals.

The average drift velocity $\overline{W}$ of an electron cloud can now be found in terms of u_1 from equations (1), (2) and (7) for different values of E_1 from 1 ev. to 6 ev. The values of the mobility coefficient h thus calculated are given in table 2, together with the corresponding values of the constant k required for equation (5).

Table 2

E in ev.	1	2	3	4	5
h	0·838	0·846	0·859	0·893	0·924
k	3·01	2·88	2·70	2·46	2·23

The r.m.s. velocity of agitation of the electrons u_1 can also be expressed in terms of Z/ρ from equations (4) and (8) by using these values of k.

It can reasonably be assumed that the coefficient g in equation (6) has the value 2·66 for all the distributions mentioned in table 1, except for the infinitely narrow one, when the value is 2. The value 2·66 is exact for a Maxwellian distribution, and the values for the other distributions cannot be very different, as these do not differ much from the Maxwellian form.

The r.m.s. value u_1 can be eliminated, using equations (2), (5) and (6). Thus

$$\overline{W} = h \,.\, (eZ/m\rho Q_0) \,.\, 1/u_1,$$

and from equation (5)

$$mu_1^2/2 = eZ/\rho Q_0 (k\lambda)^{1/2},$$

so that

$$\begin{aligned}\overline{W} &= h\{(k\lambda)^{1/2}/2\}^{1/2} \,.\, (eZ/m\rho Q_0)^{1/2} \\ &= h\{(kg)^{1/2}/2\}^{1/2} \,.\, \{(eZ/m\rho Q_0) \,.\, (m/M)^{1/2}\}^{1/2}. \qquad \ldots\ldots (9)\end{aligned}$$

This is the expression of the electron mobility in helium in terms of the atomic constants.

Consider now particular values of the mean electron energy of agitation E_1. When E_1 is approximately 4 ev. equation (5) shows that Z/ρ is 2·2 volts/cm./mm. Hg and g is 2·66; then table 2 gives $h = 0{\cdot}893$ and $k = 2{\cdot}46$. Substituting these values in equation (9) gives

$$\overline{W} = 1{\cdot}008\{(eZ/m\rho Q_0) \,.\, (m/M)^{1/2}\}^{1/2}.$$

It is of interest to note that for this particular case when E_1 is 4 ev. the numerical constant in this expression, 1·008, differs very little from the value 0·98 given by Compton's formula (1923), which did not take into consideration the probable electron energy distribution nor the variations of the atomic cross-section with the electron velocity. This is due to the fact that the effect of the electron distribution is practically cancelled out by the effects of the variable cross-section, as far as the coefficient h is concerned.

Substituting the numerical values of the atomic constants:

$$e/m = 5{\cdot}28 \times 10^{17} \text{ e.s.u./gm.},$$
$$Z/\rho = 2{\cdot}00/300 \text{ e.s.u./cm./mm. Hg},$$
$$Q_0 = 18{\cdot}8 \text{ cm.}^{-1} \text{ after Normand},$$
$$(m/M)^{1/2} = 1/85{\cdot}95;$$

then

$$\overline{W} = 1{\cdot}55 \times 10^6 \text{ cm./sec. when } Z/\rho \text{ is } 2{\cdot}2 \text{ v./cm./mm.}$$

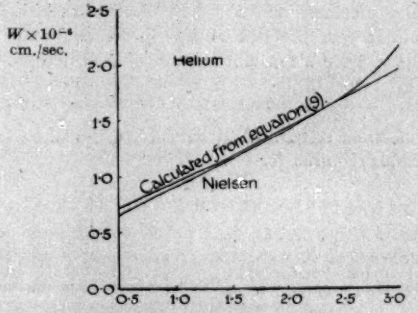


Figure 2.

Similarly, when E_1 is 2 ev. equation (5) shows that Z/ρ is nearly 1·2 v./cm./mm. Table 2 gives $h = 0{\cdot}846$ and $k = 2{\cdot}88$, so that

$$\overline{W} = 0{\cdot}996\{(eZ/m\rho Q_0).(m/M)^{1/2}\}^{1/2}$$
$$= 1{\cdot}14 \times 10^6 \text{ cm./sec. when } Z/\rho \text{ is } 1{\cdot}2 \text{ v./cm./mm.}$$

In figure 2 the values of $\overline{W}$ calculated in this way are given as a function of Z/ρ, together with the values determined experimentally by Nielsen (1936) for values of Z/ρ between 1 and 3. Over this range the agreement between the two curves is reasonably good. At values of $Z/\rho > 3$ the experimental values exceed those calculated from equation (9). This is to be expected, for the mean energy of agitation of the electron now exceeds 6·0 ev., and in this range it is not easy to obtain accurate experimental data about the electron distribution. For high mean energies the increasing importance of the losses due to excitation and ionization tends to reduce the proportion of high-energy electrons and to produce in consequence a greater proportion of low-energy electrons. These electron

groups of low energy have a correspondingly high mobility, as equation (2′) shows, and the mean drift velocity is consequently increased. In other words, when Z/ρ exceeds 3 the rate of increase of E_1 and u_1 with Z/ρ diminishes; consequently the mobility begins to increase with Z/ρ at a rate faster than $(Z/\rho)^{1/2}$, quite apart from any change in the numerical coefficient h due to the variation of Q with E. However, the agreement between the two curves in figure 2 can be regarded as evidence of the validity of the distribution formula of equation (2) over the lower-energy ranges not very different from the mean energy. The formula has already been found reasonably to account for the optical properties of uniform columns in helium over the higher energy ranges. Hence equation (1) is in accord with experimental data over the entire energy range in the régime of elastic collisions, i.e. $0{\cdot}5 < Z/\rho < 3{\cdot}0$. Thus the motions of electrons in helium may be accounted for over the full energy range by a distribution formula which represents the deviation from the ideal distribution curves produced by the variation of atomic cross-section with electron velocity.

It is interesting to note that, in neon, Druyvesteyn (1937) also obtained closer agreement between the calculated and observed electron mobilities, when the ideal diffusion distribution was modified to take into account the variation of atomic cross-section, than when a constant atomic cross-section was assumed.

§ 4. ACKNOWLEDGMENT

I am very much indebted to Mr. W. J. R. Calvert, M.A., for checking the evaluation of the integrals required for § 3.

REFERENCES

Compton, K. T., 1923. *Phys. Rev.* **22**, 333.
Cravath, A. M., 1930. *Phys. Rev.* **36**, 248.
Druyvesteyn, M. J., 1930. *Physica*, 61.
Druyvesteyn, M. J., 1937. *Physica*, **6**, 464.
Druyvesteyn, M. J. and Penning, F. M., 1940. *Rev. Mod. Phys.* **12**, 90.
Francis, V. J. and Jenkins, H. G., 1940. *Rep. Prog. Phys.* **7**, 250.
Langevin, P., 1905. *Ann. Chim.* (*Phys.*) **5**, 245.
Llewellyn Jones, F., 1931. *Phil. Mag.* **11**, 163.
Llewellyn Jones, F., 1936. *Proc. Phys. Soc.* **48**, 513.
Loeb, L. B., 1939. *Fundamental Processes in Electrical Discharges*, chap. iv., §§ 3, 4.
MacCallum, S. P., Klatzow, L. and Keyston, J. E., 1933. *Phil. Mag.* **16**, 193.
Morse, P. M., Allis, W. P. and Lamar, E. S., 1935. *Phys. Rev.* **48**, 412.
Nielsen, R. A., 1936. *Phys. Rev.* **50**, 950.
Normand, C. E., 1930. *Phys. Rev.* **35**, 1217.
Ramsauer, C., 1921. *Ann. Phys., Lpz.*, **66**, 545.
Ramsauer, C. and Rollath, R., 1929. *Ann. Phys., Lpz.*, **3**, 536.
Townsend, J. S. E., 1936. *Phil. Mag.* **22**, 145.

THE ABSOLUTE MEASUREMENT OF LARGE MAGNETIC FIELDS BY A NULL METHOD

By H. MIKHAIL, M.Sc. and Y. L. YOUSEF, M.Sc.,
Physics Department, Faculty of Science, Fouad I University, Cairo

MS. received 24 *April* 1944

ABSTRACT. A condenser method is described for measuring large magnetic fields. It is based on the simultaneous passage through a fluxmeter of a charge equal and opposite to the charge that would flow if a search coil were removed from the field. The field is calculated in terms of a capacity, a voltage, a resistance and an area.

§ 1. INTRODUCTION

Recently, Nettleton and Sugden (1939) described a null method for measuring large magnetic fields by the use of a standard mutual inductance. The underlying principle is to balance the e.m.f. produced in a search coil as it cuts the lines of force of the magnetic field by the e.m.f. induced in the secondary of a mutual inductance when the primary current is broken at the same time by the motion of the search coil, which is connected in series with the secondary. We shall give here the outline of a method in which the discharge from a condenser counterbalances the effect of cutting the magnetic field by a search coil.

§ 2. EXPERIMENTAL ARRANGEMENT

With reference to the figure, L is a search coil supported on an ebonite frame M carrying a stout wire D connected to the battery B and dipping into a mercury cup K to a depth of about 1 cm. (the optimum depth is easily found by experience). C is a standard mica-condenser, F is a fluxmeter shunted by a resistance R having the same order of magnitude as the resistance of F; P and P′ are two large resistance boxes forming a Rayleigh's potentiometer. W is a manganin wire of resistance 1 ohm, constituting a part of P, whilst E is a standard cadmium cell connected to a galvanometer G and used for the voltage calibration of the potentiometer.

§ 3. PRINCIPLE OF THE METHOD

When the search coil is suddenly raised from the field, the battery circuit is broken at K. The condenser, being connected to a closed circuit, is thus discharged, the quantity passing through F being $CVR/(R+F)$ microcoulombs, where C is the capacity in microfarads, V the voltage across P, F the resistance of the fluxmeter and R the resistance across F, including that of L. At the same time, F will receive another impulse, $HA/100(R+F)$ microcoulombs, due to cutting the field H (oersted) by the search coil, whose effective area is A cm.^{2}

With the battery B connected the right way, and with P and P′ properly adjusted, the two quantities can be made equal and opposite, so that F will show no deflection. In this case,

$$H = 100CVR/A \text{ oersted.}$$

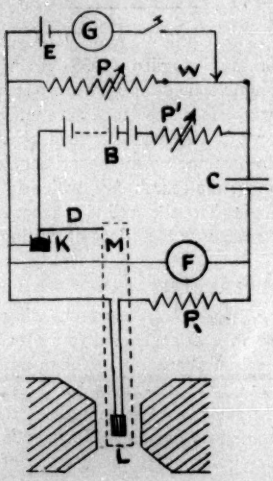
C, V and R can be known to a very high degree of accuracy. A may be determined from the dimensions of the search coil or else by an induction method due to Nettleton and Sugden (*loc. cit.*). Thus, H can be measured with accuracy.

When the balance conditions are approached, the final adjustment may be effected by slight alterations in the value of R.

§4. RESULT

The fluxmeter we used has a resistance of 99 ohms and a period of 12 seconds. The battery B has a constant e.m.f. of 30 to 100 v. The standard cell has an e.m.f. $E = 1{\cdot}0184$ v.

In a typical result, C is 2 microfarads, P is 9593 ohms, P' is 1518 ohms, R is 201·8 ohms. The part of the potentiometer which balances the e.m.f. of



the standard cell has a resistance $S = 245{\cdot}7$ ohms. Thus the charging voltage $V = EP/S = 39{\cdot}73$ v. The area $A = 124{\cdot}6$ sq. cm. The value of the field corresponding to these data is 12830 oersteds.

The method is very satisfactory and reliable. The balance is quite definite. There is no quivering of the fluxmeter, as in the method of Nettleton and Sugden, in which we observed violent quivering of the pointer. The necessary materials are easily available in every laboratory. The accuracy of the result depends on the accuracy with which the standards are known.

REFERENCE

NETTLETON, H. and SUGDEN, S., 1939. *Proc. Roy. Soc.* A, **173**, 314.

A MEAN SCOTOPIC VISIBILITY CURVE

By W. S. STILES and T. SMITH, F.R.S.,
The National Physical Laboratory

MS. received 3 *March* 1944

§1. INTRODUCTION

In practical problems concerned with vision at very low brightnesses it is useful to know the relative "intensities" of lights evaluated with respect to their effects on the scotopic mechanism of the eye (the retinal rods). The desired relative scotopic intensities can be calculated from the spectral energy distributions of the lights concerned, once the scotopic visibility curve has been fixed. This curve, which specifies the reciprocals of the relative energies of different wave-lengths required to produce a given visual effect on the rods, has been determined experimentally in several investigations. It is unlikely that the final conclusions in practical problems will be greatly modified if one of these experimental curves is used rather than another. However, the differences between the various curves are considerable, and the general adoption of a mean curve may help to remove fictitious differences between the results of different investigations and may facilitate the framing of specifications. Such a mean curve was prepared for a Committee of the Air Ministry, and a rather wider circulation now seems desirable.

The scotopic visibility curve S against λ given in the accompanying table was derived as a weighted mean from the results of five investigations:—

(1) S. Hecht and R. E. Williams (1922).
Method: brightness matching at about $2{\cdot}7 \times$ threshold intensity. Matching field (contrast type) of 22° diam. Monocular vision. 48 sets of measurements. Number of observers not stated.

(2) W. S. Stiles and B. H. Crawford (1934).
(Mean curve given as P_λ in "Visibility of Light Signals", R. & M. No. 1793. Aero. Res. Com. 1937, p. 38.)
Method: absolute threshold for 5° parafoveal vision. Test spot, $0°{\cdot}7 \times 0°{\cdot}16$, exposure 1 sec. Monocular vision. Two observers.

(3) K. S. Weaver (1937).
Method: brightness matching at brightness $6{\cdot}5 \times 10^{-5}$ e.f.c. (white light). Central spot (2°·6) in 14° surround. Natural pupils (binocular vision). Fourteen observers.

(4) W. S. Stiles (1939).
Method: absolute threshold for 5° parafoveal vision. Test spot, $1° \times 1°$, exposure 0·06 sec. Monocular vision. One observer.

(5) H. V. Walters and W. D. Wright (1943).
Method: brightness matching in an area 10° from the fovea, at a brightness corresponding to 5×10^{-4} photons of equi-energy white light (equivalent to 8×10^{-6} e.f.c.). Monocular vision. Two observers

Weight 3 has been given to the results of investigations (1) and (3) and weight 1 to the others. The quantity meaned is log S. The final values of S have been smoothed and multiplied by a factor to give S a maximum value of unity. None of the experimental investigations extends to wave-lengths below 410 or above 750 mμ, but a limited extrapolation has been made to enable a definite scotopic value to be calculated for filters transmitting in the ends of the spectrum. It is sometimes convenient to have a precisely defined value in such cases, even though, with present knowledge, the validity of the result cannot be assessed. At the short wave-length end account has been taken of a value given by Goodeve, Lythgoe and Schneider (1942) for a young eye at wave-length 365 mμ.

Mean scotopic visibility curve (S). Smoothed values

λ (mμ)	S	$\log_{10} S$	λ (mμ)	S	$\log_{10} S$
370	0·0007	$\bar{4}$·846	590	0·0934	$\bar{2}$·970
380	0·0026	$\bar{3}$·415	600	0·05437	$\bar{2}$·735
390	0·0074	$\bar{3}$·869	610	0·03121	$\bar{2}$·494
400	0·0177	$\bar{2}$·248	620	0·01763	$\bar{2}$·246
410	0·0373	$\bar{2}$·572	630	0·009802	$\bar{3}$·991
420	0·0713	$\bar{2}$·853	640	0·005362	$\bar{3}$·729
430	0·1253	$\bar{1}$·098	650	0·002887	$\bar{3}$·460
440	0·2047	$\bar{1}$·311	660	0·001533	$\bar{3}$·186
450	0·3121	$\bar{1}$·494	670	0·0008023	$\bar{4}$·904
460	0·4450	$\bar{1}$·648	680	0·0004144	$\bar{4}$·617
470	0·5948	$\bar{1}$·774	690	0·0002115	$\bar{4}$·325
480	0·7451	$\bar{1}$·872	700	0·0001067	$\bar{4}$·028
490	0·8757	$\bar{1}$·942	710	0·0000535	$\bar{5}$·728
500	0·9663	$\bar{1}$·985	720	0·0000267	$\bar{5}$·427
510	1·0000	0	730	0·0000133	$\bar{5}$·124
520	0·9668	$\bar{1}$·985	740	0·0000066	$\bar{6}$·820
530	0·8677	$\bar{1}$·938	750	0·0000033	$\bar{6}$·515
540	0·7183	$\bar{1}$·856	760	0·0000016	$\bar{6}$·209
550	0·5487	$\bar{1}$·739	770	0·0000008	$\bar{7}$·902
560	0·3881	$\bar{1}$·589	780	0·0000004	$\bar{7}$·594
570	0·2552	$\bar{1}$·407	790	0·0000002	$\bar{7}$·285
580	0·1576	$\bar{1}$·198			

§ 2. SCOTOPIC VALUES AND SCOTOPIC TRANSMISSION

If a coloured light has relative spectral energy distribution $Ed\lambda$, the ratio of its scotopic to its ordinary photometric intensity may be defined as

$$\sigma = \int ESd\lambda \Big/ \int EVd\lambda,$$

where V is the standard visibility function of the eye. Given this ratio for light of a particular energy distribution, the scotopic value of candle-power, brightness, etc., is obtained by multiplying the corresponding photometric magnitude by σ. This definition of scotopic values fixes the units so that for monochromatic light of wave-length 530 (approx.) (at which wave-length $S = V$) scotopic and photometric values are identical. If, on the other hand, scotopic units are chosen so that scotopic and photopic values are identical for white light of colour temperature 2360° K., the appropriate values of σ are given by

$$\sigma = 2{\cdot}020 \int ESd\lambda \Big/ \int EVd\lambda.$$

This particular choice of units is of interest as feeble brightnesses are now usually specified by comparison with a white at 2360° K.

§3. COMMENT ON THE PROPOSED MEAN SCOTOPIC VISIBILITY

Not all the available scotopic curves have been used in deriving the mean curve. The early determinations of König and Ritter and Abney and Watson for example, have been omitted. In normal times it would not be justifiable to propose a mean curve before improved measurements on many more eyes had been made. The proposed curve, however, is imperfect also in another way. The rods of the dark-adapted eye are more than 100 times as sensitive as the cones for wave-lengths less than about 510 mμ. In moving to the red end of the spectrum, the difference diminishes, and in the red the cones may be the more sensitive. Thus although the measured parafoveal threshold values in most of the spectrum certainly depend only on the rods, in the red both rods and cones may be contributing to the response.* If so, the observed threshold in the red would be higher were the rods alone acting, and the values of S determined by parafoveal threshold measurements will be too high in the red. Similar difficulties in an aggravated form arise in determinations of S by brightness-matching methods. It is probable that the sensitivities of the parafoveal rods and cones are not widely different in the red, and that the glossing over of the difficulty just raised does not result in an error large compared with errors from other sources.

It might be argued that, after all, what are required in applications are the resultant sensitivities of the dark-adapted eye, the sensitivity of the rods (or cones) for the colours to which the rods (or cones) are more sensitive, and the experimentally determined "mixed" sensitivity when both rods and cones are contributing to the response. This view is complicated by the different relative contributions of rods and cones in different retinal regions (at the fovea, for example, there is no rod response), and by differences in the integrative properties of rod and cone systems. Moreover, in some cases (for example, when considering the relative disturbing effects of red and white light on rod vision) the objects to be seen are feebly illuminated in white light and are certainly seen by rod vision, although the disturbing light may provide a predominantly cone response.

An alternative method of determining S, appropriate in this case, is provided by an extension of the threshold method. The parafoveal threshold value for a test patch of white light (or better, monochromatic green $\lambda = 510$) is determined for zero background brightness. The threshold so determined will be the rod threshold. Keeping the test-patch colour unchanged, the intensities of uniform monochromatic background brightnesses necessary to raise the measured threshold by a given factor, say 5, are determined in turn. For each monochromatic background the reciprocal of its intensity in energy units obtained in this way defines the power of light of that colour to raise by the

* The relative contributions of the rods and cones to the threshold in the red will no doubt depend on the angular area and exposure time of the test patch. This is to be expected because of the different integrative properties of the rod and cone systems.

given factor the threshold of the rods or, in other words, to destroy to a given extent the dark adaptation of the rods. In one investigation (4) in which essentially this method was used, the results, as far as they go, show that the reciprocal of the background intensity in energy units varies with λ in the same way as S. It appears, in fact, that the powers of lights of different colours to stimulate the rods are approximately proportional to their powers to destroy the dark adaptation of the rods, a result which we should expect to be true on almost any theory of the mechanism of rod vision. The merit of this method of determining S is that it can be extended as far as desired at the red end of the spectrum, while remaining free from the uncertainties concerning the possible participation of the cones, which embarrass other methods. The few results obtained by the indirect method confirm the earlier conclusion that the proposed mean scotopic visibility curve is not gravely in error in the red.

§4. MODIFICATION OF THE VISIBILITY CURVE WITH INCREASE IN THE BRIGHTNESS LEVEL

It is not possible to specify a brightness level (photometric or scotopic) below which the retina, excluding the rod-free fovea, behaves to light of all colours as though it contained only rods. There is reason for thinking that even in the extreme case of zero background brightness, the parafoveal threshold in the red is dependent on both the rod and cone responses, their relative importance varying with the characteristics of the test patch. The resulting difficulty in determining the true scotopic (i.e. the rod) visibility curve in the red has already been noted. It was decided, however, to accept the relative visibilities obtained at zero background brightness (threshold method) or at a very low field brightness (matching method) in deriving the mean scotopic visibility curve.

When the background brightness is not zero, the visibility function may still be defined as the reciprocal of the parafoveal threshold in energy units for a monochromatic test patch applied as an additional stimulus to the background. It is known that a white background brightness below about 3×10^{-6} e.f.c. has no effect on the threshold sensitivities of either rods or cones, so that for that condition the visibility curve will be the same as in total darkness. As the background brightness is increased from 3×10^{-6} to 3×10^{-3} e.f.c. the rod threshold for any particular stimulus wave-length is raised progressively, while the cone threshold is unchanged. The observed parafoveal threshold for wave-lengths in the red cannot be much changed because it depends, in part at least, on the cone response, and the cone threshold is unchanged. In the rest of the spectrum, however, the observed threshold coincides with the rod threshold (the cone threshold being very much larger), and will be raised in the same proportion for all wave-lengths. Thus the visibility curve is distorted by a relative rise of the values in the red, the shape of the rest of the curve being unchanged, with its maximum remaining at about 510 mμ. With increase in the background brightness above 3×10^{-3} e.f.c. the cone threshold begins to rise, but the rod threshold is increasing more rapidly, so that finally the rod exceeds the cone threshold at every wave-length and a pure cone visibility curve is observed.

If the background brightness is red, its scotopic value will be much less than that of a white of the same photometric intensity. Suppose, for example, its scotopic value is one-tenth that of the white. Then the rod threshold will be unaffected by a red background brightness less than $10 \times 3 \times 10^{-6} = 3 \times 10^{-5}$ e.f.c., and below this value the scotopic visibility curve will be unmodified. With higher red background brightnesses, the visibility curve will tend to change in the same way as with a white adapting brightness, except that a greater photometric brightness of red will be required to produce the same modification of the curve. There is some evidence to show that quite high brightnesses of red light of the order of 100 e.f.c. may be insufficient to complete the transformation of the visibility curve to a pure cone curve.*

Wright's determinations (5) of the visibility curve by brightness matching in the parafovea show that when the brightness of the matching field is increased from a very low value, the progressive distortion of the visibility curve is similar to that described above.

It is apparent from this discussion that no simple rule can be given to determine the relative visual effects of lights of different colours in all conditions. A further complication, ignored above, is the fact that the cones form a triple system and, under certain conditions of observation, the cone visibility curve is distorted by changes in the relative contributions of the three component cone systems.

REFERENCES

GOODEVE, C. F., LYTHGOE, R. J. and SCHNEIDER, E. E., 1942. *Proc. Roy. Soc.* B, **130**, 380.
HECHT, S. and WILLIAMS, R. E., 1922. *J. Gen. Physiol.* **5**, 1.
STILES, W. S., 1939. *Proc. Roy. Soc.* B, **127**, 64.
STILES, W. S. and CRAWFORD, B. H., 1934. *Proc. Roy. Soc.* B, **113**, 496.
WALTERS, H. V. and WRIGHT, W. D., 1943. *Proc. Roy. Soc.* B, **131**, 340.
WEAVER, K. S., 1937. *J. Opt. Soc. Amer.* **27**, 36.

* The experiments in question were made with a test spot exposed in brief flashes. Rather different results might be obtained with a continuously exposed test patch.

ON ELECTRICAL RESONANCE

By C. H. COLLIE,
Clarendon Laboratory, Oxford

MS. received 23 *December* 1943; *in revised form* 24 *May* 1944

ABSTRACT.—Resonance in a coaxial line and a dipole aerial is treated by an elementary method based upon damped waves. An attempt is made to estimate the error due to the use of simplifying assumptions.

§ 1. INTRODUCTION

OWING to the importance of the radio industry, many physics students are studying experimentally the phenomena of resonance in electrical circuits. Not only are they carrying their studies beyond the stage reached by any except the most advanced text-books, but they are able to devote much less time than heretofore to mastering the theoretical background of their experimental work. In the circumstances it seemed worth while to find out

how far it was possible to cover the salient experimental results by elementary methods. No great novelty is claimed for the results which follow, which have doubtless been used for many years by individual teachers. It is, however, hoped that their collection, using a common notation and treated from a single point of view, will serve a useful purpose in the present circumstances.

By an elementary treatment is meant one which makes the maximum possible use of the student's existing knowledge and experience and introduces the minimum of new material ; it will be different for students of different previous training.

The present treatment is intended for men hitherto trained as physicists who but for the war would have completed their undergraduate course without any quantitative study of radio problems. Such men should be well acquainted with the general properties of waves, and the treatment given makes as much use as possible of the concept of the plane wave.

This departure from convention is most noticeable in the treatment of the resonant aerial, which is treated essentially as a problem in the diffraction of a plane wave. Naturally those who have spent much time in mastering the theory of transmission lines will prefer a treatment in which as close an analogy as possible is drawn with the terminated transmission line. The results obtained by either method are, of course, the same to the first order, i.e. to the only order of accuracy to which they can lay any claim.

Finally it may be useful to draw attention to a third approach, due to Schelkunoff, based upon a somewhat simplified boundary-problem treatment. His method has the advantage of obtaining the results due to Page and Adams and King with the use of more elementary functions. It will appeal to mathematicians, but not, I believe, to those whose talents are physical insight rather than mathematical dexterity. To this numerous class the boundary-condition method, with its concentration on the steady state to the complete exclusion of the preceding events in which the steady state was built up, often wears an artificial and unconvincing aspect.

§ 2. THE THREE TYPES OF CIRCUIT

Electrical systems may be classified according to whether they are large or small compared with the wave-length of the radio waves at the resonant frequency and according to whether the energy is supplied to the circuit from the external source over a region large or small compared with this same wave-length.

Class 1. *All components small compared with the resonant wave-length.*

This class is represented by an LC circuit with lumped constants, and need not be treated here. The only point which calls for comment is that students are usually a little hazy about the nature of the resistance R which governs the decrement of the circuit, often attributing its high value to the skin effect. The simple skin effect is, of course, too small to have much effect in ordinary coils. The high damping is due to currents induced in the neighbouring turns (proximity effect), * and can well be represented by a sheet of copper foil placed in the magnetic field of the solenoid.

* Tables of the proximity effect are to be found in the Smithsonian Physical Tables. (A full treatment is given by Butterworth, 1921.)

Teaching experience shows that a thorough understanding of the physical reason for the connection between decrement and sharpness of tuning (Q) is an essential foundation for subsequent work.

Class 2. The circuit comparable with the wave-length and energized locally.

As an example of this class of circuit, consider the quarter-wave coaxial line (figure 1) excited by the field from the small probe P near the open end. The electromagnetic wave radiated from the probe is received by the line and trapped in it, since it is unable to lose energy.*

The electrical state of the line is thus to be found by tracing the history of the wave which has been continuously injected into the line from the probe.

If the radial field E of the wave is represented by the damped wave, $E = E_0 e^{-\alpha x} \cos 2\pi(\nu t - x/\lambda) =$ real part of $E_0 e^{j\omega t - px}$, where $\omega = 2\pi\nu$, $p =$ the propagation constant $\alpha + 2\pi j/\lambda = \alpha + j\beta$, then, allowing for a phase change of π

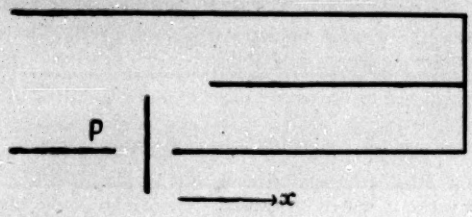


Figure 1.

when the wave is reflected at the closed end, the waves travelling from left to right on the quarter-wave line are

$$E_0 e^{j\omega t - px}, \quad -E_0 e^{j\omega t - p(x+2l)}, \quad E_0 e^{j\omega t - p(x+4l)} \ldots,$$

while those travelling from right to left are

$$E_0 e^{j\omega t - 2pl + px - j\pi}, \quad -E_0 e^{j\omega t - 4pl + px - 2j\pi} \ldots.$$

Thus the total electric force at any point on the central conductor is given by

$$E_0 e^{j\omega t - px}(1 - e^{-2pl} + e^{-4pl} \ldots) + E_0 e^{j\omega t + px}(-e^{-2pl} + e^{-4pl} - e^{-6pl} \ldots)$$
$$= E_0 e^{j\omega t} \cdot \frac{e^{-px} - e^{p(x-2l)}}{1 + e^{-2pl}} = E_0 e^{j\omega t} \cdot \frac{\sinh p(l-x)}{\cosh pl}.$$

We are most interested in the behaviour of this expression near resonance when $l = \lambda/4 \pm \Delta l$.

Defining the departure from resonance ϕ by the equation

$$2\beta l = \pi \pm \frac{4\pi\Delta l}{\lambda} = \pi + \phi,$$

the expression for the electric field becomes

$$E_0 \cdot e^{j\omega} \cdot \frac{2\sinh p(l-x)}{e^{pl}(1 + e^{-2pl})} = E_0 e^{j\omega t} \cdot \frac{2\sinh p(l-x)}{je^{pl} \cdot e^{j\phi/2} \cdot (1 + e^{-2pl})}.$$

* The outer member of the coaxial line must have a diameter less than the critical diameter for a T.M. wave. This condition is usually amply satisfied.

On expanding the numerator and neglecting the difference between $\cos\phi/2$, $\cosh\alpha.(l-x)$, and unity, one obtains

$$\sqrt{2}.E_0.e^{j(\omega t-\phi/2-\eta)}\frac{\{\cos\beta x+\sin\phi/2\sin\beta x+j\sinh\alpha(l-x)(\sin\phi/2\cos\beta x-\sin\beta x)\}}{(\cosh 2\alpha l-\cos\phi)^{\frac{1}{2}}}$$

The phase angle η is given by $\tan\eta=\dfrac{2\phi}{\phi^2+4\alpha l}$ and the common phase factor $e^{-j(\eta+\phi/2)}$ mainly determines the distribution of field on the line.

It will be observed that when $l=\lambda/4$ exactly, and $\phi=0$, the form of the wave is very nearly $\cos\beta x$, and that the out-of-phase component is zero at both extremities. There is no value of ϕ for which the component in quadrature is everywhere zero.

The electric field at the open end is given by

$$E_0e^{j\omega t}.\frac{\sinh 2\alpha l-j\sin\phi}{2.(\sinh^2\alpha l+\sin^2\phi/2)}.$$

At resonance, $\phi=0$ and the voltage amplification is $1/\alpha l$. The response curve, which shows how the electric field at the open end depends upon the degree of mistuning ϕ, is given to the second order by the expression $\left\{1+\dfrac{\sin^2\phi/2}{\sinh^2\alpha l}\right\}^{-\frac{1}{2}}$.

The electric field lags behind the phase of the incident wave by $\tan^{-1}\dfrac{\sin\phi}{\sinh 2\alpha l}$

It appears probable on general grounds that resonance will be maintained as the exciting probe is moved towards the closed end of the line, but it is not easy to devise a formal proof of this. The method which has been given is only applicable to excitation which is localized at the ends of the line.

This restriction is common to all methods of calculation which do not assume an actual discontinuity at the point where the energy enters the line, e.g. a conducting tap to another line. The difficulty is a physical one, since it is not possible to devise a method of feeding in the energy, which is confined to a very small length of the line and which does not at the same time have a large but incalculable distorting effect upon the waves flowing up and down the line.

Numerical values

For a metre wave and a coaxial line whose inner and outer diameters are 1 and 5 cm. respectively, α* is $8{\cdot}96.10^{-6}$, so the voltage magnification is $4{\cdot}4.10^3$ and the phase shift when the line is 1 mm. off tune is 88°. This remarkably sharp tuning illustrates the great superiority of lines over circuits with lumped components. Experimental evidence that the predicted sharpness of tuning can be approached in practice will be found in a paper by Willshaw (1940), which contains an exhaustive treatment of the tapped resonant line.

Class 3. *The circuit is comparable with the wave-length and the electric force is applied throughout the circuit.*

Consider a thin wire with its axis placed parallel to the uniform electric field of a plane electromagnetic wave.

Since the radius a of the wire is small compared with the wave-length of the

* For the necessary formulae see Terman (1937).

incident radiation, it casts no shadow and the electric field acts uniformly upon its surface, inducing a uniform current in it. One would thus expect a long thin wire to have a uniform current in it and to form the axis of an outward, radially symmetrical, cylindrical wave. The scattered energy would be very small because, when the wire is thin, it acts as an inductance and the current is very nearly wattless.

This is indeed the case so long as the wire is infinitely long and the conditions at the end of the wire do not have to be considered. If the wire is finite the current flowing into the free end of the wire produces an intense local electric field which gives rise to a travelling wave on the wire * in a way essentially similar to the wave produced in the coaxial line considered in the previous section.

Let the wire OA (figure 2) be l cm. long, the electric force parallel to it be $E_0 e^{j\omega t}$ and the induced uniform current $-ji_0 e^{j\omega t}$. The free ends of the wire must be current nodes and so a current wave $ji_0 e^{j\omega t - px}$ will start from 0 and a

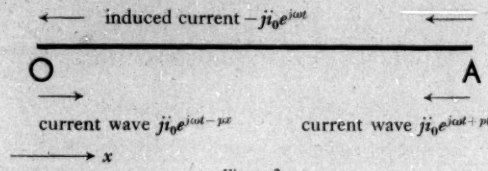


Figure 2.

current wave $ji_0 e^{j\omega t + p(x-l)}$ from A. The amplitude i_0 of these waves is left to be determined later.

Then, as before, the total wave-current in the wire is

$$ji_0e^{j\omega t}\{e^{-px}(1-e^{-pl}+e^{-2pl}-e^{-3pl}\ldots)+e^{px}(e^{-pl}-e^{-2pl}+e^{-3pl}\ldots)\}$$
$$=ji_0e^{j\omega t}\cdot\frac{e^{-px}+e^{p(x-l)}}{1+e^{-pl}}.$$

It is convenient to refer this expression to an origin at the centre of the wire, i.e. put $y=x-l/2$. The current is then

$$\frac{ji_0\cosh py}{\cosh pl/2}\cdot e^{j\omega t},$$

and exhibits the phenomenon of resonance when the length of the wire is in the neighbourhood of a whole number of half-wave-lengths. At resonance, the current due to the wave greatly exceeds the small uniform current which initiated it, and the resonant aerial can extract and scatter energy from a much larger area of wave-front than corresponds to the cross-section of the wire.

The harmonic current in the wire can be represented as the product of an amplitude factor $f(\alpha l)$ depending on the length of the wire, and a position factor $\cosh py=\cosh \alpha y\cos \beta y+j\sinh \alpha y\,.\sin \beta y$, which determines the distribution of current along the wire. When l is nearly a resonant length (i.e. $\frac{n\lambda}{2}\pm\Delta l$),

* A formal justification for the assumption of waves on a single wire has been given by White, 1932.

the amplitude factor is $\{\sinh^2 \alpha l/2 \cdot \cos^2 \theta + \cosh^2 \alpha l/2 \sin^2 \theta\}^{-\frac{1}{2}}$, with θ the degree of mistuning, given by $\theta = \pi \Delta l/\lambda$ radians.

Thus at resonance ($\theta = 0$) the main part of the current is in phase with the electric force of the incident wave and has an approximately cosinusoidal distribution ; the small out-of-phase component flowing at the extremities is just sufficient to balance the uniform inductive component and make the end of the wire a true current node in agreement with the initial assumptions.

The change of phase ψ near resonance is given by $\tan \psi = -\coth \alpha l/2 \cdot \tan \theta$. The magnitude of the current at resonance, hitherto left arbitrary, can be calculated from considerations of energy. The electric field of the wave is doing work on the element of current between y and $y+dy$ at the mean rate of $Ei/2 \cdot \cos \beta y \cdot dy$, so that the mean power absorbed by a dipole is

$$\int_{-\lambda/4}^{+\lambda/4} \frac{Ei}{2} \cos \beta y \cdot dy = \frac{Ei\lambda}{2\pi} \text{ watts},$$

if E is measured in volts/cm. and i in amperes.

But it is known from the theory of radiation resistance that the current, if sinusoidal, is dissipating energy as radiation at the same rate as would a resistance of 73 ohms at the mid-point of the dipole. Hence

$$\frac{73i^2}{2} = \frac{Ei\lambda}{2\pi} \quad \text{and} \quad i = \frac{E\lambda}{73\pi} \text{ amp.}$$

Similarly, the effect of radiation damping can be represented, as far as conservation of energy is concerned, by a resistance uniformly distributed at the rate of 292 ohms per wave-length, and thus α (the damping constant of the wave) for a dipole aerial is given by

$$\alpha = \frac{r}{2Z_0} = \frac{292}{2\lambda Z_0} = \frac{146}{\lambda Z_0} \text{ cm.}^{-1}.$$

It is to be observed that α is not independent of the length of the aerial ; for an aerial many half-wave-lengths long, the radiation resistance and, therefore, αl, only increase very slowly (logarithmically) with l. Explicitly, the radiation resistance of an aerial m half-wave-lengths long is given by Stratton as $72{\cdot}45 + 30 \ln m - 30 \operatorname{Ci}(2\pi m)$. Values of the cosine integral $\operatorname{Ci}(x) = -\int_x^\infty \frac{\cos t}{t}\, dt$ are given in Jahnke-Emde ; contributions to the radiation resistance from this term are small and can be neglected for values of m exceeding 4. Thus the amplitude factor $f(\alpha l)$ hardly depends upon the length of the aerial, and the current in a resonant aerial is approximately independent of its length. Thus the main difference between a long and a short resonant aerial is the direction in which the energy abstracted from the wave-front is re-radiated.

Before the value of α can be calculated, Z_0, the surge impedance of a free wire, must be known. This is most easily approximated to by calculating C, the capacity per unit length of a wire of length l and radius a, from the expression $C = 1/2 \ln l/a$ given by Jeans (1925) and Abraham (1932).

Hence Z_0 for a dipole aerial of diameter d is $138 \log_{10} \lambda/d$ ohm, and α, the damping constant for the electric waves upon it, is

$$\frac{146}{Z_0 \lambda} = \frac{1{\cdot}057}{\lambda \log_{10} \lambda/d}.$$

Finally, it is possible to make some estimate of the end correction of a dipole aerial. We have hitherto assumed that the free end was a true current node of zero capacity, so that the phase change on reflection was zero for a voltage wave and π for a current wave. Actually the free end will behave as if it had a small capacity C, and will, therefore, not be a true node. Since from line theory a capacity C is equivalent to a length of open-ended line δl given by $\frac{1}{\omega Z_0} \cdot \tan \frac{2\pi \delta l}{\lambda} = C$, it is clear that the resonant length observed in practice will be shorter than $\lambda/2$ by $2\delta l$.

An approximate value of C may be calculated (1) by assuming the aerial to end in hemispherical caps whose capacity will approximately be $d/4$ cm., on which assumption the resonant length would be about $\lambda/2 - d(1 + 2{\cdot}3 \log_{10} \lambda/d)$, or (2) from the expression d/π for the capacity of a flat disc of diameter d and negligible thickness. It should be noted that the end correction due to the finite current flowing into the end of the dipole is in addition to that usually given (King, 1927; Schelkunoff, 1941), which has its origin in the decrease in inductance of the aerial near its ends. That both corrections will have the same sign seems likely, since they both decrease the value of Z_0, the surge impedance of the aerial near its extremities.

Numerical values for a dipole

Taking, as before, $\lambda = 100$ cm., $d = 1$ mm., we find that $\alpha = \frac{1{\cdot}05}{\lambda \log_{10} \lambda/d} = 3{\cdot}5 \times 10^{-3}$ cm.$^{-1}$. The resonant current is $\pi/4$ out of phase when $\tan\theta = \tanh \alpha l/2$, and so $Q = \lambda/4\Delta l = 9{\cdot}0$. Similarly the end correction for the dipole would be 6·5 mm. = 1·4 %.

§ 3. DISCUSSION

The chief attraction of the treatment which has been outlined is that it enables the main phenomena of resonance to be treated by a single elementary method. All the difficulties inherent in a strict treatment based on Maxwell's theory are segregated by making use of Hertz's well known calculation of the radiation resistance of a dipole. Experience shows that students who find great difficulty in following Hertz's arguments can with profit make use of the idea of radiation resistance. It is believed that the results obtained are nearly true, but the direct experimental evidence is meagre.

The weakness of this method of obtaining the approximate current distribution in an aerial lies in the fact that it does not inherently lead to an estimate of the magnitude of the possible error. However, some idea of the error involved can be obtained by comparing the value of the resonant current i, adjusted to give the correct dissipation of energy, with the value calculated from the known value of i_0, the current induced in an infinite aerial. Nicholson (1910) has shown that the inductive current flowing in such an aerial of diameter d is given by

$$\frac{2j\lambda E}{\pi d} \cdot \frac{Y_0(\pi d/\lambda)}{Y_0^2(\pi d/\lambda) + \pi^2 J_0^2(\pi d/\lambda)},$$

where the Y and J are Bessel functions of the second and first kinds. Substituting this value for i_0, one can readily show that for small values of d/λ the

resonant current in the centre of the dipole is $\frac{E\lambda}{72{\cdot}95\pi}$ amp. compared with $\frac{E\lambda}{73\pi}$ amp. found in the previous paragraph. The close agreement indicates that for very fine wires the approximate theory will be correct; the necessary value of d/λ is about 10^{-7}, a factor considerably smaller than is usually met with. For the aerial for which numerical values have been calculated ($d/\lambda = 10^{-9}$) an idea of the error involved is seen by noting that the phase angle of i_0 calculated from Nicholson's results differs from $-90°$ by about 17°. Actually the formulae developed represent the experimental results much more accurately than this, because, as is easily shown by graphical integration, minute changes in the distribution of the current produce large changes in the component of the electric field parallel to the axis of the aerial. Hence a current distribution which satisfies Maxwell's equations for a very thin wire will also very nearly satisfy them for a thicker wire. The best experimental evidence available for this statement is the general agreement between the values of radiation resistance obtained directly from transmitting aerials and indirectly from receiving aerials. This agreement will only be obtained if the current distribution in the two aerials is similar. However, the current distribution in the receiving aerial gives rise to a uniform tangential field just sufficient to neutralize the uniform field of the incident wave.

The current distribution in a transmitting aerial gives a zero tangential field over the surface of the aerial, since the source of power is imagined to be concentrated at its mid-point. The effective identity of the two currents can only be reconciled with the radically different boundary conditions if small differences in current distribution cause large changes in the electric field parallel to the axis of the wire, as previously stated.

§ 4. ACKNOWLEDGMENTS

I have to thank Dr. Roaf for some helpful discussions, and Mr. Keeley for his interest and encouragement.

REFERENCES

Abraham and Becker, 1932. *Classical Theory of Electricity and Magnetism* (London: Blackie), p. 64.
Butterworth, 1921. *Phil. Trans.* A, **222**, 51.
Fowle, 1924. *Smithsonian Physical Tables* (Washington City).
Jaknke and Emde, 1928. *Funktionentafeln* (Berlin: Teubner), 2nd edition.
Jeans, 1925. *Electricity and Magnetism* (Cambridge: The University Press), p. 248.
King, 1937. *Philos. Trans.* A, **236**, 381.
Nicholson, 1910. *Proc. Lond. Math. Soc.* **11**, 104.
Schelkunoff, 1941. *Proc. Inst. Radio Engrs.* **29**, 495.
Stratton, 1941. *Electrodynamics* (London and New York: McGraw Hill), p. 444.
Terman, 1937. *Radio Engineering* (London and New York: McGraw Hill), p. 699.
White, 1932. *Proc. Camb. Phil. Soc.* **18**, 356.
Willshaw, 1940. *Phil. Mag.* **29**, 572.

PHYSICS AND THE SCIENTIFIC INSTRUMENT INDUSTRY

By A. J. PHILPOT, O.B.E.,
British Scientific Instrument Research Association, London

Lecture delivered 3 *March* 1944

THE association of physics with scientific instruments is so general and occurs over so wide a field that a real difficulty arises as to which aspect of this subject can most profitably be considered within the relatively short time at our disposal. I have thought it best, therefore, to base my remarks upon practical experience gained through a connection lasting for nearly twenty-five years with the British Scientific Instrument Research Association, and to indicate at least the minimum requirement which the scientific instrument industry has for physics and for physicists in order that it may be maintained in a state of health. The subject cannot very well be explored in the absence of any definition of what constitutes a scientific instrument, and, unfortunately, no definition exists which is completely satisfactory to all parties concerned with scientific instruments and their manufacture. It might, however, be agreed that an instrument becomes a scientific instrument when its manufacture cannot be achieved by a knowledge of, and skill in, arts and crafts alone, and when the production of the instrument and an understanding of the principles of its operation require an extensive knowledge of physical principles.

Every scientific instrument incorporates a physical principle and has been developed either as an application of a particular advancement in physical knowledge or as an implement to facilitate particular lines of investigations or of measurements in the realm of science. The experimental physicist is perhaps apt to take too narrow a view of the scope of scientific instruments and to think of them as being either instruments specially designed to assist in the teaching of physics or very specialized instruments which enable him to observe certain phenomena in connection with his research activities and to measure effects of these phenomena precisely. These types of instruments are the tools of the experimental physicist, and not unnaturally his greatest interest has been in such instruments. It must not be forgotten, however, that there is a very large range of instruments used by the civil population, by industry, on land and sea and in the air, which are also in every sense scientific instruments, and which incorporate physical principles which have been established as the result of experimental work carried out by the physicist throughout the years. It is these instruments which, in the eyes of the general public, provide an obvious monument to the work of the physicist, and it is perhaps unfortunate that an equal interest has not been shown by physicists in those instruments of a utilitarian character which are used daily, often by non-scientific people, as in those instruments which are more closely and more immediately connected with his academic and experimental activities.

It is particularly appropriate at the present time that the attention of physicists should be drawn to their responsibilities in connection with all classes of scientific instruments. The need for the greater application of science to all departments of the nation's life has been widely proclaimed and generally accepted. Not only does the scientific instrument provide in itself the most immediate application of science to industrial and every endeavour, but no general application of science can be made without an extensive use of scientific instruments. It would be disastrous if industry, for example, were generally to adopt scientific control of its processes, with a consequent large-scale installation of scientific instruments, without a corresponding incursion of physicists and physics into this field of endeavour. If an extensive use of scientific instruments in industry demands that there shall exist in industry adequate personnel having a knowledge of physics, how necessary it is that the scientific instrument industry should be permeated with a knowledge of physics, for only thus can the application of science in industry, as represented by the general use of scientific instruments, be as complete and as efficient as possible. If science is to be applied, it is evident that it must be applied by scientists, and in the general impact of science on the life of the nation the work of such scientists is complementary to that of the pure experimental scientist. Any attitude of mind which discriminates between these two types of scientists is to be deplored. In other realms it has long been accepted that the fact that Paul planted whilst Apollos watered provided no criterion for assessing the relative merits of these two worthies. There are, however, grounds for the belief that there has existed in some degree a wall of partition between pure and applied science which has hindered the efficient application of science to everyday activities and which in particular has had a devitalizing effect upon the scientific instrument industry.

There is no essential difference between the qualities and qualifications of successful experimental physicists in the pure and applied fields. It is probably true that the " applied physicist ", through his contact with spheres of activity outside the laboratory, has a broader outlook in certain directions than has the pure physicist. On the other hand, the extra commitments of the applied physicist make it difficult for him to keep completely abreast with the trends of modern physics. Essentially, however, the successful experimental physicist, of either type, is a man to whom physics is a living subject. The man to whom physics appears as a heap of dry bones, the precise dimensions of which are to be carefully measured and memorized, may acquire an apparent knowledge of physics which may even serve him well in the examination room; but if such a man enters the field of research his efforts will only be successful in so far as his work is skilfully directed from above rather than from his own original thought. A man, however, who sees in the dry bones the possibilities of articulation, mentally clothes them with flesh, and gives to them life and movement, has the makings of the true experimental physicist. A simple illustration is provided by the physics of the bending of bars. To one student the phenomenon of the bending of a bar is associated in his mind solely with the obtaining of that set of formulae, the numerical values of the denominators of which vary with the methods of support and loading. Another student, however, visualizes the state of strain in the bent bar, and any suggestion of imposing

further stresses on the system would immediately lead him to think of the possibility of disaster. The first student would be solely concerned with the effect of such imposed stresses upon his formulae. The truth of the above will be generally recognized, but the point to be observed is that the true experimental physicist, whether in a University laboratory, in a Research Association, or in Industry, is a man who not only knows his physics but has an understanding of the subject and tends automatically to apply his knowledge and understanding when faced with any physical problem. All experimental physicists follow the road of application, although certain details of the road may vary in different regions.

In considering the application of physics to the scientific instrument industry, one's thoughts naturally turn to those spectacular applications which occur from time to time—some with relative suddenness and others very gradually over a long period. One need only mention the development of the electron microscope or the almost infinite variety of uses made of the phenomena associated with the emission of electrons, and which has resulted in the wide incorporation of thermionic valves, photo-electric cells and cathode-ray tubes in modern instruments. Such applications of physics to scientific instruments are, however, being constantly brought to the notice of this Society, and possibly, in the minds of many, represent the main association of physics with the scientific instrument industry. It is of course a very important association, but it is not proposed here to do more than merely mention it. For there is another application of physics to the scientific instrument industry which in the long run is probably more vitally important to the industry than these spectacular ones which receive so much popular attention. It is highly desirable that avenues should exist whereby all modern trends in physics can influence and make their impact upon scientific instrument development; but it will be very difficult to find such avenues if the scientific instrument industry does not include a very large number of men having a really sound and live knowledge of those fundamental principles of physics which still provide the pillars upon which the whole structure is supported. A knowledge of wave mechanics may, or rather will, have a long-term influence on scientific instrument developments. A knowledge of classical mechanics is an immediate and vital necessity. The point can best be illustrated by making mention of the type of knowledge which must be available to the manufacturer of certain classes of instruments.

Every graduate in physics has some knowledge of optics and is acquainted with the principles of operation of the classical types of optical instruments. He probably, however, has very little idea of the peculiar and intensive application of the principles of optics which is made by the designer in the computation of an optical instrument system. (In passing it may be mentioned that there is a certain grim humour in the facile way in which, on the lecture-room blackboard, an object arrow has its counterpart, as a result of the intersection of two or three specially chosen rays, in a perfect image arrow. The optical computer may well spend his life-time attempting to achieve this ideal result.) The elimination of aberrational defects, both spherical and chromatic, and of traces of coma, the consideration of the image-forming properties of skew rays, the obtaining of the flattest possible field, and many other matters which might be mentioned,

all of which are somewhat lightly touched upon in the normal optics course, do in fact constitute a whole system of specialized optics within the general optical framework. The work of the optical computer has never received the recognition due to it, and it is not too much to say that the subject of applied optics is worthy of consideration by the best mathematical physicists. The connection between the work of the optical computer and physics is an obvious one, but in the production of an optical instrument the application of physics by no means ends with the optical computation. In fact, the whole work of the optical computer may be vitiated if the optical instrument be constructed in the absence of a general knowledge of physics. The computer must assume that the optical material he is using—in general optical glass—is optically homogeneous, and at one time the final performance of an instrument provided the only criterion as to whether this was so. In recent years the phenomenon of interference, as utilized in the interferometer, has been employed to provide a tool for the ready testing of optical materials for homogeneity. Further, an absolute assessment of the performance of a lens system, such as a microscope object glass, can be obtained by the use of an interferometer with special adaptations, and it is now no longer necessary to have recourse to the personal judgment of an expert microscopist. The knowledge of the optical instrument maker must not, however, be confined to optics. An optical instrument may, and probably will, be subject to marked changes of temperature. In such circumstances an enclosed instrument may well act as a still, and any volatile constituent of lubricating greases, sealing compounds, and paints used in the instrument may be deposited as a film on the optical components, with a consequent impairment of performance. Changes of temperature will result also in differential expansion between the glass components and the metal mounts and between the varying types of glass in a compound lens. The quality of the instrument may be much lowered at elevated temperatures if the instrument has been constructed in the absence of a knowledge and a ready appreciation of the effects of thermal expansion. The lubricant used in the moving parts of an optical instrument may develop, under moist conditions, acid properties, and in such circumstances all the conditions favourable to the formation of a voltaic cell and to the development of electrolytic currents will exist. All the troubles of corrosion may well ensue. It is quite possible that the relative merits of two instruments, of equal optical merit, are largely determined by the extent of the knowledge of physics which has formed a background to their construction.

The moving-coil galvanometer is an historical instrument and its action is based on firmly established physical principles, but this does not mean that the problem of the construction of a highly sensitive galvanometer is by now merely a mechanical one. Its construction does demand a high degree of mechanical skill, but its ultimate useful sensitivity may well be determined by physical actions other than the main ones controlling the movement of the coil. If the materials of which the coil is made exhibit any ferromagnetism, the holding of the coil in a deflected position will cause the coil to acquire a magnetic moment of some permanence, and the existence of this will in turn prevent the coil from returning to its original zero position. If the materials be paramagnetic or diamagnetic, the coil will tend to orient itself in relation to the direction of greatest

magnetic field, so that the sensitivity of the instrument may very well vary, failing absolute uniformity of field, with the extent of deflection. In addition, the movement and accuracy of reading of a highly sensitive galvanometer can easily be affected by faulty elastic properties in the very fine suspension, by the existence of parasitic thermo-electric currents set up as a result of temperature variation, and by the creation of electrolytic currents resulting from the action of moisture on varnishes and other insulating materials. Somewhere, therefore, in the production of a satisfactory highly-sensitive galvanometer there must exist a knowledge of physics other than that relating to the main principles of operation of the instrument.

Interest in the level, with its moving bubble, is largely restricted to its more domestic uses on the carpenter's bench or the billiard table, and it is perhaps not generally realized that the highly sensitive level, capable of responding to seconds of arc, is, and has to be, a scientific instrument of precision. The skill of the glass grinder in producing the requisite accuracy of curve of long radius inside the level tube is greatly to be admired, and that skill is essential to the successful construction of a highly sensitive " bubble " level. The level tube, however, has to be mounted, and, for use on certain types of surveying instruments, for example, has to be suitably housed. Now the shape and length of the bubble in a level are partially determined by surface-tension affects, and if an accurate reading is to be made by taking the mean of the positions of the two ends of the bubble, it is obviously necessary that no variation in temperature along the length of the bubble shall cause differences to exist in the values of the surface tension at the two ends of the bubble. As a consequence, every effort must be made to make the housing of the level as nearly as possible a constant-temperature enclosure. In the choice of design and of materials, considerations of heat radiation, of emissivity and absorption, and of heat conduction must arise, and again it appears that the production of an apparently mechanical instrument demands a knowledge of physics over and above efficient craftsmanship.

The compass is an ancient instrument operating according to laws known to antiquity. A large number of modern compasses are not of the " dry " type, but are instruments in which the magnet with its attached compass card moves in a liquid contained in a liquid-tight compass bowl. The liquid is chosen of such viscosity as will allow the rate of motion of the magnetic system to be most suited to the particular conditions of use. In the compass is a variety of metals of which the bowl itself, the pivot, the pivot-holder, the solder, and other parts are composed. All of these are immersed in a liquid, and should such a compass be constructed in the absence of any knowledge of contact differences of potential and of the laws of electrolytic action, disaster may well ensue. The compass, instead of serving as a safeguard and a help to the mariner or airman, may well be used as a showcase for the demonstration of the physics of corrosion. An instance was brought to the notice of B.S.I.R.A. a few years ago of a certain type of liquid compass suddenly developing marked corrosion after a long history of satisfactory use. It was shown that the occurrence of this defect was solely due to the employment in a minute joint of a solder which differed slightly in composition from that previously employed in the trouble-free compasses. An

innocuous potential system had thus been converted into a vicious one. The craftsman's skill and the knowledge that one pole of a magnet tends to point to the north are not sufficient for the manufacture of a sound liquid compass. A much wider knowledge of physics must be behind the manufacture if the many pitfalls are to be avoided.

The instances given above have been taken almost at random from a memory of practical experience. They could be multiplied a hundred or a thousandfold. It is to be hoped, however, that they will have sufficed to show that, whilst it is necessary for those connected with scientific instrument manufacture to have at least an acquaintance with modern trends in physics, it is essential that the scientific instrument industry should be permeated with a knowledge of the well-established and, one might almost say, old-fashioned fundamentals of physics. It is perhaps permissible to wonder whether the present system of acquiring a knowledge of physics is best suited for the provision by our universities of men capable of vitalizing the scientific instrument industry. A generation or so ago a general, but elementary, grounding in physics was given in schools. The real knowledge of the subject was acquired in the more mature atmosphere of the university. To-day the schoolboy acquires a knowledge of physics almost equivalent to that required for a pass degree a generation ago. He proceeds to the university and almost immediately is instructed in the more erudite aspects of the subject. After graduation, he may, under the direction of his professor, carry out researches with the object of advancing physical knowledge. It would be presumptuous, and foolish, generally to condemn a system which has produced signal advances in knowledge in recent years, but, purely from the point of view of the requirements of the industry, it may be noted that much of that knowledge of physics with which a healthy industry must be permeated is acquired under the restrictive conditions of inadequate practical facilities and of limitation of apparatus common to the average school. There is naturally a fascination and an excitement in the acquiring of modern atomic and nuclear physics, and one can understand the student's concentration of interest in such subjects. It would be unfortunate, however, from the point of view under consideration, if this interest were accompanied by any lessening of appreciation of the vital importance of the more prosaic physics of yesterday.

Our knowledge of physical principles is almost entirely due to the work of the experimental physicist throughout the years. Scientific instruments are based on these principles and thus, in a real sense, the physicist is the parent of the scientific instrument. Parenthood carries responsibilities, and these children of physics should not be left solely to the care of foster parents. One thinks wistfully of the time when Newton was not only interested in the formulation of the laws of optics, but was much concerned with the processing of lenses and prisms. The names of some of our most famous instrument firms indicate that, up to the beginning of this century, leading physicists were actively engaged in the directing of instrument manufacture. Might it not at least be of advantage if examples of modern applied scientific instruments were found in our physical laboratories? A student gains a knowledge of thermo-electric phenomena by experimenting with an inexpensive couple of iron and copper; surely it would be excellent if there were available to him a modern industrial thermo-electric

recording pyrometer, the principles of operation of which he might study and the accuracy of which he might determine by comparison with the laboratory resistance-thermometer. The value of a theoretical knowledge of the microscope might well be supplemented, or rather complemented, by the possibility of using a microscope of good performance, which instrument, it may be remarked in passing, is most admirably suited for the demonstration of the practical implications of diffraction phenomena. The atmosphere of our university physical laboratories may have tended to become too much that of the purest physics. One would hesitate to ask for a change of outlook in the academic experimental physicist, but a plea might be put forward for some widening of his interests. When one remembers that the close association which existed between the universities and the scientific instrument industry, until relatively recent times, coincided with the period of greatest prestige of British instruments, one cannot but hope that a revival of interest in the industry will be manifest in the near future, to the advantage of the physicist in the better provision of the instrumental tools so necessary to his work and to the manufacturer in the more ready access to that knowledge which is the life-blood of his industry.

DISCUSSION

Mr. F. TWYMAN. Instruments may be described as scientific for several reasons. Some are so called because science is peculiarly necessary for their design and manufacture ; examples are binoculars, gun-sights, galvanometers, barometers. Other instruments are used for the scientific control of industrial processes, as in the pyrometric control of the heat treatment of metals.

Again, reproductions, for educational purposes, of apparatus with which notable additions to physical or chemical knowledge were made in the past are quite reasonably called scientific instruments ; so are new instruments devised for their own use by research workers.

Now while, as Mr. Philpot has shown, a knowledge of physics—and particularly of classical physics—is required in our industry for solving problems of manufacture, the other aspects to which I have referred bring other occasions for the scientist, and in particular the physicist, to be of use.

First comes the business of finding new markets for our existing products. Marketing in our industry requires something more than the simple devices employed in the advertisement of cigarettes, beer, or cosmetics, and the selection of reliable travellers or selling agents in our own or foreign countries. It demands qualifications which are only possessed by the trained scientist.

Each firm should have one or more physicists or physical chemists whose business it is to keep acquainted with the manufacturing processes employed in industry, and the methods of measurement used by chemists, biologists and so forth, so that he can talk to the technicians in their own language. He should seek in this way, and by reading the technical journals of the world, for applications in industry of the instruments which we make.

The second function of such a man is to look for fresh instruments to make. The instruments made by scientific men for their own researches form a rich field for exploitation; every month such instruments are described in the scientific journals of the world. If the research has yielded important results there are hundreds of other laboratories in the world that will want copies of them, while some of them will ultimately find a use in industry.

Some manufacturers deplore the growing tendency of college and institutional laboratories to manufacture instruments for original researches instead of coming to us for them. But how foolish to object to other people inventing things for us to make without cost to us !

Thirdly, the physicist might select some of the most important of the instruments used in classical experiments of the past, so that copies could be made available for educational purposes.

The manufacturer may say, " Where is the money to come from ? " There is a simple and adequate answer. Charge it to advertisement account. An average British Board of Directors will then be quite happy about it.

DISCUSSION

on paper by J. T. Kendall (*Proc. Phys. Soc.* **56**, 123 (1944)) entitled "The rectifying property of carborundum".

Dr. A. FAIRWEATHER. In referring to some earlier work of mine on the same topic Mr. Kendall asserts (p. 123) that " Fairweather .... omits to mention with what type of crystal he is dealing ". On pp. 514–516 of my paper it was pointed out that there are five known crystalline modifications of carborundum, and that the crystals employed in my work consisted mainly of the second modification. The method of classification adopted by Mr. Kendall, i.e. according to colour, as either green or black or blue, does not seem very definite. The point at issue is, therefore, whether colour or crystal structure is the significant factor, and whether they are alternative descriptions of the same thing. There are several possibilities : thus, different colours might be associated with—

(*a*) different crystalline forms of pure carborundum;

(*b*) different forms containing the same impurity;

(*c*) different impurities in the same form;

(*d*) different impurities in different forms, a specific impurity being, perhaps, always associated with the same crystalline form.

Secondly, Mr. Kendall's fusing technique for the preparation of a low-resistance contact (p. 124) is not free from objection. With the particular material for which it was devised, viz. carborundum, it involved raising the temperature of the crystal to " about 1200° c." Without an experimental check, it cannot be accepted that the properties of the crystal and/or its surface remained unaffected by this treatment. For other materials such a procedure would be impracticable, and the technique would seem, therefore, to be of much more limited application than that described in my paper.

Lastly (p. 128), Mr. Kendall mentions that " At low voltages ..... the contact resistance is linear. Over a certain voltage range non-linearity is observed, and at high voltages the resistance apparently tends to become linear again. The latter may, however, be a *contact breakdown* effect ". A high voltage ohmic region obtained as a result of a contact breakdown (i.e. a non-reversible change in contact structure) cannot be regarded as a fundamental property of a particular contact structure. But there is a high-voltage ohmic region which is not a breakdown effect : the complete current/voltage characteristic, including this high-voltage ohmic region, is reproducible provided that overheating at high voltages is avoided by ensuring that the test current flows for a sufficiently brief period of time. *Prolonged* current flow in this region may, however, result in a non-reversible change in contact structure. The characteristic regions were identified, and their significance discussed, in my paper (p. 500 *et seq.*) : the work described therein related specifically to smooth reversible phenomena (p. 499, p. 509 *et seq.*), and particular care was taken to avoid confusion with non-reproducible effects.

AUTHOR'S reply. The work described in my paper has shown that there are two types of crystal to be found in commercial carborundum, one behaving as an *excess* and the other as a *defect* semi-conductor. This is in confirmation of earlier work, e.g. Dowsett (1921), Gokhberg (1937), which has not, apparently, been generally appreciated. The author has been unable to correlate these electrical properties with crystal structure. As far as present experiments have proceeded, *all* commercial carborundum (including crystals of

both electrical types) is of structure-type II. This evidence is solely from powder photographs and is not quite conclusive, since it is not easy by this method to distinguish between the various structures. It is hoped at a later date to obtain rotation photographs of crystals of different electrical properties.

The various structures are all very similar, and are based on a structure-unit of four carbon atoms arranged tetrahedrally about a central silicon atom. The carbon tetrahedra are joined by sharing corner atoms, and different sizes of unit cell are formed according to the relative rotation of adjacent tetrahedra about their common axis—which becomes an axis of *screw*-symmetry in the crystal. It does not seem likely that such small differences in structure are the cause of such a striking difference as *excess* and *defect* electronic conduction. It remains a possibility that different impurities may cause the production of different structural types as well as different electrical types, but there is no necessity for a correlation between the two.

The author's separation of carborundum crystals according to colour appears to be much more satisfactorily related to electrical properties. The separation between green on the one hand and blue and black crystals on the other is quite definite, and very easily made visually ; it corresponds exactly with the separation according to electrical type. Moreover, it is well known that the formation of colour-centres in crystals is a phenomenon associated with the presence of free electrons, so that one should not be surprised by this correlation of colour and electrical properties.

Commercial carborundum is undoubtedly an *impurity* semi-conductor, and it is to these impurities that its unusual electrical properties must be ascribed. Unfortunately the nature of the impurities is not yet known ; they may be carbon or silicon in excess of stoichiometric proportions, i.e. vacant lattice points, or they may be foreign atoms, e.g. Fe, Al, etc., either substitutional or occupying interstitial positions in the crystal lattice. Since some carborundum crystals behave as *excess* and others as *defect* semi-conductors, it would be convenient to believe that the first possibility is correct, and that excess silicon or carbon results respectively in *excess* or *defect* conductivity, but this view can only be taken as conjectural. Further electrical work starting with pure silicon carbide, and with impurities under exact control, is obviously necessary.

With reference to Dr. Fairweather's second paragraph, the well-known potential-probe method used by him is certainly of more general application than the technique of fusing a contact on to a crystal surface. In the case of carborundum this drastic operation did not, however, affect the electrical properties of the crystal. This was shown by fusing two carbon rods to a crystal in the usual way, and then breaking off the rods at the contact surface, when the electrical properties were quite unaltered. It will be realized that the crystal was at 1200° c. for only a second or two. If it is kept in air at that temperaturre for several minutes a film of silica of high electrical resistance is formed at the surface.

This fusion technique was adopted because of the difficulty of obtaining *single* carborundum crystals of a size sufficiently big to use the potential-probe method. It was considered important to exclude possible electrical effects due to the inter-crystalline contacts which are obviously present in carborundum fragments of appreciable size.

With reference to Dr. Fairweather's third paragraph, the tendency towards ohmic resistance of carborundum contacts at high voltages is explained by him (1942) on the assumption that the resistance of the contact becomes small compared with the bulk resistance of the crystal. This may be true, but the evidence is not yet complete, and the experimental difficulties of obtaining " smooth reversible phenomena " at the higher voltages have not yet been overcome. It is hoped to investigate this range at a later date by applying impulse voltages of very short duration to single carborundum crystals, and observing the current-voltage characteristic directly with a high-speed cathode-ray oscillograph.

CONTACT LENSES

A discussion held by the Optical Group, 18 February 1944

Mr. K. CLIFFORD HALL. (1) Contact lenses are frequently referred to as a new idea, and it therefore comes as a surprise to many to learn that it is over 117 years since the theory of contact lenses was first conceived.

In 1827, at a time when corneal astigmatism was just beginning to be understood, Sir John Herschel wrote about the possibilities of correcting this defect by means of a small glass lens worn on the eye itself and filled with gelatine.

Sixty years later, in 1887, an artificial-eye maker named F. E. Müller, of Wiesbaden, had a patient whose only remaining eye had had the eyelids removed because of a cancerous growth. Müller blew a thin glass shell to protect the eye. It is reported that the wearer retained his sight and wore this glass until his death, twenty years later. So this, the first contact glass, was for *protection* only.

In the following year Dr. A. E. Fick, of Zürich, made the first attempt to improve vision by means of a contact glass. It was his belief that if a spherical glass filled with a liquid were fitted over an eye suffering from keratoconus, the distorted vision would be neutralized. His experiments proved that this was so, and for the first time a satisfactory means of restoring vision in conical corneas was found. Unfortunately at this stage the wearers could not tolerate the lenses in their eyes very long at a time.

In 1889, August Müller, an ophthalmologist of Kiel, worked out the necessary specifications for a contact lens to correct his own 14 diopters of myopia. Himmler, a Berlin optician, ground a contact lens to these requirements, and Müller found that all but 0·50 diopters of his myopia was neutralized. This was probably the first contact lens to correct ametropia. It was a disappointment to Müller to find that he could not wear the lens for more than half an hour.

About the same time, Kalt, of Paris, attempted to *cure* conical cornea by pressure with contact glasses.

By the beginning of this century, contact lens literature was centred round the question of the superiority of blown or ground contact lenses. It was generally agreed that blown lenses were more comfortable, but that ground lenses were more accurate. Because of this, a number of attempts were made to combine the two types. At least three contact lens workers sent batches of blown contact glasses to Zeiss's, requesting them to grind the corneal portion. Zeiss broke them all.

By about 1920 Zeiss improved their ground contact lenses. They used a harder glass and obtained a better polish. Then they introduced a refinement whereby the transition between the corneal and scleral curves was rounded off. These improvements resulted in greater comfort, and the Zeiss ground contact lens became very much more popular.

In 1930, Professor Heine, of Kiel, introduced a Zeiss fitting set of afocal contact lenses with a wide range of varying scleral and corneal radii. The set was expensive, but great things were expected of it. Soon, however, it was discovered that only a very few eyes could be satisfactorily fitted with it.

In 1935, the firm of Müller-Welt, of Stüttgart, succeeded in grinding the optical correction on blown contact lenses. The firm of Danz, of New York City, accomplished the same feat at about the same time.

Since then progress has been rapid, and I cannot here deal with it in detail. Sufficient to say that recently we have seen the introduction of moulded contact lenses in both glass and plastic. Contact lenses tinted in the entire corneal portion, or just the pupillary area, or around the iridic area only, are all possible today. It is also possible to incorporate prismatic and cylindrical powers in contact lenses.

(2) Among the advantages of contact lenses over ordinary spectacles are that they provide a complete, unrestricted field of vision. Moreover, as the wearer is constantly looking through the optical centre of his contact lenses, he is not bothered by distortion and aberration. In high myopia he is rewarded by better visual acuity. This is because

the contact lens provides an increased size of retinal image. For instance, with 6 D. of myopia there is an increase of 10 per cent, and with 20 D. of myopia an increase of 50 per cent.

In anisometropia—the condition where there is a marked inequality between the refraction of the two eyes—the spectacle-wearer has difficulty in maintaining single vision whenever his eyes move away from the optical centres of his lenses. Contact lenses will, of course, eliminate the effect of this prismatic displacement, as the wearer is constantly looking through the optical centres of his correcting lenses.

Another factor in anisometropia is the unequal-sized retinal images which result. Contact lenses alone cannot rectify this ; in fact they will intensify the difference. However, with contact lenses and auxiliary spectacles even large differences can frequently be equalized. Briefly, the procedure is to over-correct one eye and under-correct the other with the contact lenses, and then with spectacles correct the residual error.

In keratoconus, ordinary spectacles are of very little help, whereas contact lenses will neutralize all the surface irregularities, thereby providing the wearer with really useful vision, often even restoring a normal acuity. Corneas which are scarred and faceted as a result of injury, mustard-gas burns, ulceration, and so on, can also be benefited in the same way.

The value of contact lenses to sportsmen is considerable. Contact lenses cannot fall or be knocked off. The wearer can swim and dive in them with safety ; and they will not be affected by rain, mist or spray. The contact lenses, occupying a position protected by the orbital bones and the eyelids, are less liable to breakages than spectacles. This already minimized danger can be completely eliminated by the use of the new flexible plastic contact lenses.

Contact lenses are practically invisible when in use. This is the factor which appeals to most wearers, and is particularly useful in certain occupations. Here is a decided psychological value to those who feel they are disfigured by the wearing of spectacles.

Contact lenses are greatly appreciated by aphakic patients. These, by the way, are unusually tolerant to them, generally speaking. Post-cataract spectacles are almost always ugly, so that the aphakic contact-lens wearer benefits both by vision free from distortion and by an improved appearance.

In monocular aphakia combined with emmetropia, single binocular vision is impossible with spectacles, for two reasons, viz. (*a*) an induced muscle imbalance occurs as the eyes are turned. Approximately one prism diopter of error is created for every millimetre the visual axes are directed away from the optical centres ; (*b*) the two retinal images will differ in size by as much as 33 per cent. With contact lenses the induced muscle imbalance is completely eliminated because the lenses ride on the eyes ; and the unequal images, by the reduction in the vertex distance, are reduced to a minimum—only 9 per cent.

(3) So much for some of the advantages of contact lenses. If one is to consider them fairly, it is only right that their disadvantages and limitations should be mentioned. Only in this way is it possible to assess the probable future of these lenses.

Contact lenses are expensive. This is due to the considerable time required for fitting each patient. The practitioner's equipment is likely to be costly. Moreover, the lenses themselves require enormous skill in manufacture, and the process does not lend itself to methods of mass-production. It would seem that, however great the demand for contact lenses should become, they will always cost very much more than ordinary glasses.

The future of contact lenses is also limited by psychological difficulties. Many people are nauseated by the thought of putting a lens in the eye. The contact-lens patient needs to be keen to be successful ; almost always perseverance is necessary at first. It is unfortunate that there are some people who are unable to wear perfectly-fitting contact lenses without discomfort, due to the fact that their eyes are hypersensitive. Because of this, it is necessary for prospective wearers to test their tolerance with trial glasses or unfinished shells before going to the expense of manufacturing the optically-worked contact lenses.

The big unsettled problem is a means of eradicating a haziness of vision which generally occurs after the lenses have been worn for some hours. This is sometimes referred to as *Fick's Phenomenon* or *Sattler's Veil*. The mistiness is accompanied by the appearance of faint-coloured halos noticed round lights ; the phenomena persist for a short while even after the lenses are removed. A slit-lamp corneal microscope examination reveals a

softening of corneal epithelium. Many theories as to the causation have been put forward, but no remedy for all cases has been found. Amongst the possibilities which have been suggested as the cause are interference in the temperature of the eye and osmosis. My own approach to the problem is to have each patient try a number of different solutions for the fluid lens. These solutions are all adjusted to a *p*H of about 8·2, and vary in salinity. The patient generally finds that one particular strength gives better results than all others ; he is advised to continue with this, and frequently discovers that longer and longer periods without haziness will result. A number of patients would seem to have completely eliminated the phenomena. One patient wore his lenses for 24 days without removal, and another wore hers for five days. Neither of these cases experienced the haziness, although both were troubled for the first few months of wearing the lenses. Incidentally, it is very unusual for patients to wear their contact lenses for such very long periods.

In spite of these limitations and difficulties, numerous cases have been very satisfactorily fitted with contact lenses. Many patients suffering from subnormal vision, incapable of improvement by other means, have regained normal visual acuity with contact lenses. Others have been able to enjoy a renewed self-confidence by being able to dispense with unsightly spectacles. Still others have been able to undertake work denied to spectacle-wearers.

Sgt. HENDEN. I think it would be helpful if Mr. Hall would discuss the advantages of each type of contact lens in turn, i.e. the ground glass lens, the glass and plastic lens, the moulded plastic lens and the ground plastic lens.

Mr. KEELER. I found in fitting cases of keratoconus with contact glasses that it is necessary to clean the apex of the cornea from the inside surface of the contact glass to prevent the possibility of causing an opacity.

Mr. H. V. WALTERS. Could Mr. Hall tell us whether, in cases of high myopia, there is likely to be any discomfort due to the high edge-thickness of the corneal portion of the contact glass, resulting, possibly, in rather a steep slope to the scleral portion ?

Mr. E. F. FINCHAM. The correction of refractive errors by means of spectacle lenses can be entirely satisfactory only in those cases where the errors are approximately regular spherical or sphero-cylindrical deviation from the normal, although frequently, in cases of irregular refraction, spectacles do appreciably improve vision.

In those cases where an irregularity occurs in the refractive index or in the form of an internal surface such as the crystalline lens, no optical device can give complete correction, but when disease or damage has left the cornea with an irregular surface a contact lens attached to the eye can restore the normal optical form. It was for the treatment of such cases that the contact lens was originally proposed by Herschel (1827) and was employed by A. E. Fick (1887).

The purpose of this short note is to give some particulars of the results of treating a case of conical cornea with contact lenses and to show photographs of an illuminated section of the cornea without the lens and with the lens in position.

Conical cornea is a condition in which the centre of the cornea bulges forward, due, it is thought, to a congenital weakness in that area. It is generally stated that in these cases the centre of the cornea is abnormally thin, but in the photographs I am showing it will be seen that whereas the thickness is approximately normal at the centre, it is considerably above normal at the periphery.

The patient's vision without glasses was very low. With the right eye she could count fingers at 1 metre, the left eye gave vision of 6/60.

The best spectacle prescription was :—

R.E.	−14·0 D sph.	which gave 6/60
	−7·0 D cyl. ax. 180	
L.E.	−4·50 D sph.	which gave 6/24 poor.
	4·00 D cyl. ax. 170	

Contact lenses were prescribed by Mr. P. M. Moffatt, of the Royal Westminster Ophthalmic Hospital. He advised that the lenses should be made to touch the cornea in the centre with the object of trying to retard the progress of the deformity.

The contact lenses were fitted and made by Mr. A. E. Plaice, of Messrs. Clement Clarke, Limited. The power of the right lens was +1·75 D. and the left +2·75 D. The inside radius of the corneal segment was 8·0 mm. in each case. With these lenses the patient obtained 6/9 vision with each eye and 6/6 binocularly. She wore them all day without discomfort.

The photographs were made about a year after the lenses were fitted. The slit-lamp method of illumination is very suitable for showing the general form of the corneal surface in any given meridian. For this purpose the image of an illuminated vertical slit was projected on to the eye and the camera was placed in the same horizontal plane and at an angle of 70° with the axis of illumination. It was found that the polished surfaces of the contact lens were invisible in the slit-lamp beam; the lens was therefore coated with a thin layer of matt varnish, which caused just sufficient scattering of the light to make the surfaces apparent.

By comparison with the photograph of the normal cornea the abnormality of the form of the keratoconus is quite apparent. In this case there was an opacity in the centre of the cornea which is seen white in the photograph. This would appear to extend over

Normal cornea. Conical cornea. Conical cornea with contact lens.

more than the pupillary area, but it is probable that the bright light used in photography caused the pupil to be abnormally small. The variation in thickness of the cornea is a point of interest. I have no evidence whether the increase toward the periphery is a peculiarity of this case or whether, in spite of the customary statement to the contrary, this is the usual condition, and the slit-lamp method of studying it has given a better revelation of the true state of the cornea than did the older histological methods. It is probable that either an abnormal central thinning or a peripheral thickening would have the same effect in causing the membrane to assume a conoidal form.

The photograph of the conical cornea with the contact lens in position shows that the lens, instead of touching only the apex of the cornea, as it was originally fitted, is in contact over almost the whole of the surface ; the back surface of the lens is just visible at the top of the section where there is a small space between it and the cornea. The effect of the contact lens in this case was to effect a mechanical correction by pressing the cornea back into a spherical form.

In conclusion, I wish to express my indebtedness to Mr. P. M. Moffatt for sending me this case to photograph, and to Mr. A. E. Plaice for supplying the records of the patient's vision and particulars of the contact lens.

Mr. K. Clifford Hall. In reply to Sgt. Henden, ground glass contact lenses are favoured because of their simplicity in fitting. By using a trial set of various known spherical and toroidal curves, the best possible fit is gauged. Before going to the expense

of making the finished optically corrected contact lenses the patient's tolerance to them can be ascertained with these trial lenses.

The combined glass-plastic lens of Feinbloom seems to me to have the advantage of lending itself to mass-production methods, with an ultimate reduction in cost. If the demand justified it, it would be possible to stock the glass corneal portions in various radii and, in addition, stocks of the various plastic scleral portions, in all likely spherical and toroidal curves. In this way the complete contact lens could be made up by fixing the two portions together, in much the same way as stock spectacle frames can be glazed with stock ophthalmic lenses. Personally, I feel that the translucent scleral portion of this lens is inferior in appearance to a transparent material—though it has been claimed that a semi-opaque material does at least conceal a red and inflamed eye if the fitting has failed to come up to expectations.

Moulded plastic contact lenses have not gone much beyond the experimental stage in this country, so it is not possible for me to speak from personal experience. (Incidentally moulded glass lenses are probably used more extensively in England than anywhere else in the world, and have reached a very high standard of perfection.)

Moulded lenses are generally much larger than ground lenses, so that when made of plastic one would expect the 50 to 60 per cent reduction in weight to be a considerable advantage. The flexibility of plastic contact lenses seems to me to be a decided advantage, in that less pressure is felt when the eye changes shape with rotation. Other workers have, however, disagreed with me on this point, and are actually searching for a rigid plastic material.

Finally, ground plastic contact lenses have the advantages of ground glass lenses in that the fitting technique with them is simplified. The factors of lightness and flexibility mentioned above also apply. The fact that plastic contact lenses scratch easily is not such a disadvantage as might be supposed, because, when wetted by the tears on the outside and the saline solution on the inside, the surface defects are filled and their effect neutralized. A contact lens with a scratched scleral portion, however, does seem to be less comfortable than one which is well polished. Of course it is possible to have worn lenses repolished whenever desired. Plastic contact lenses have a slightly better light transmission than glass, though it is doubtful whether this is an advantage, as photophobia is a common trouble with contact-lens wearers. Perhaps the greatest advantage of the plastic lens is that of safety, for it is almost unbreakable. Although the risk of fracture of a glass lens in position in the eye is slight, the hazard is entirely eliminated when plastics are worn. It may be claimed that an eye is actually safer under this protection than when no correction is in use.

I am very interested to have Mr. Keeler's observations. It has been my practice also to clear the apex in keratoconus so as to avoid the risk of possible corneal scarring. In three cases I have made an exception to this, and have watched these cases over a period without noticing any ill-effects. Even if the apex of the cone should become opaque it would not be certain that the contact lens would be to blame, as it is not uncommon for this to happen in advanced keratoconus when contact lenses have never been used.

It is my experience that the visual acuity can often be considerably improved by having the contact lens press the cornea back into shape. Maybe this eliminates or reduces irregularities of the posterior corneal surface, which cannot, of course, be neutralized by the fluid lens. One frequently finds that pseudo-lenticular astigmatism is present when the contact lens clears the cornea in keratoconus, and that it can be reduced in quantity and changed in axis by having the contact lens press on the cornea.

In reply to Mr. Walters, I feel that owing to the comparatively small area of the contact lens which carries the optical correction, the thickness does not need to be very great even with powers of a considerable degree. I have fitted cases as high as 24 D. of myopia and the edge thickness has not caused discomfort. In both high myopia and aphakia it is wise to reduce the effective optical area to 10, 8, or even 6 millimetres diameter, depending upon the maximum size of the patient's pupil. By this means, in the same way as a lenticular spectacle lens, the weight and edge thickness can be very greatly reduced. Even so, with high powers the slight increase in weight tends to cause the lens to sag, so that it is necessary to allow for this in the scleral fitting by " tightening " the fit in the vertical meridian, and in certain cases the " effective optic " needs to be decentred very slightly upwards in the corneal portion to compensate for an unavoidable lag.

THE MEASUREMENT OF ABERRATIONS

A discussian held by the Optical Group, 19 *November* 1943

Mr. J. Home Dickson. *Abstract*: Methods are described for measuring the axial spherical aberration for any wave-length by visual observation of pencils of light isolated by a perforated plate placed in front of the lens. The extension of the method to oblique pencils is indicated, thus providing a method of measuring coma for all wave-lengths. Simple methods of measuring astigmatism, curvature and distortion are also described. Emphasis is placed on the importance of obtaining the visual focus and relating the results to the photographic focus at all parts of the field. Theoretical discussions are omitted from this description, which is intended only to indicate the experimental technique.

1. *Axial spherical aberration and chromatic variation of spherical aberration.* The method is a development of that described originally by Hartmann (1904), in which small pencils of light passing through zones of the lens are isolated by means of a perforated diaphragm placed in front of, or behind, the lens. In the Hartmann method the pencils form patches of light on photographic plates placed at positions inside and outside the focus. By measuring the distances between the patches on the resulting negatives the course of the rays from different zones of the lens can be traced and the spherical aberration calculated for each zone.

The original Hartmann method is somewhat laborious and slow and is limited in its application, and a visual method has been evolved which not only gives more accurate results, but gives a very clear indication of the properties of the lens to the observer. A simple further development is to use monochromatic light for the source and thus to find the spherical aberration for each wave-length.

2. The source of light used is a monochromator consisting of a Hilger constant-deviation spectroscope, in which the eyepiece is replaced by a second slit. A small diaphragm placed over the entrance or exit slit forms a pinhole source. It is necessary to focus the monochromator so that the image of the entrance slit is formed at the exit slit and the instrument is calibrated by means of a known source, so that, by rotation of the wave-length drum, monochromatic light of a known wave-length issues from the exit slit. A pointolite lamp provides an adequate source of light which is condensed by means of an auxiliary lens on to the entrance slit. For comfort, the source should be bright, since the violet and red ends of the spectrum are difficult to see. The monochromator is placed at a distance from the lens under test at least equal to twenty times the focal length and, if necessary, a good surface mirror can be used to increase the distance.

3. In front of the lens under test is placed the perforated plate used for isolating the zones of the lens. The plate must be made with care to ensure that pairs of holes are equidistant from the optical axis of the lens. In the apparatus actually used the plate was mounted on a separate stand and could be rotated in its own plane about the lens axis ; the lens itself could also be rotated about its axis. By this means it was possible to observe if the alignment of the axes of the disc and lens were coincident. The test is conclusive provided the lens does not suffer from bad figure. The holes in the perforated disc are comparatively large, and nothing is gained by making the holes small, as diffraction effects prevent good observation. Hartmann gives 1/200 of the focal length for the diameter of the holes, but in the apparatus used, the holes were $\frac{1}{4}$ inch in diameter for a 20-inch lens with the pinhole source at a distance of 30 feet. To prepare the plate, a piece of 18-gauge brass was carefully cut to a circle 6 inches in diameter on a lathe and the centre of the circle accurately marked. While still in the lathe chuck, eleven concentric circles were marked at intervals of $\frac{1}{4}$ inch radius, commencing with one of radius $\frac{1}{4}$ inch. A number of radii were then drawn and the positions of the holes were marked so that the pattern of the holes made a double whorl, as shown in figure 1. The reason for this

choice was to obtain a good separation of neighbouring holes, and so prevent excessive diffraction and interference effects, which tend to make observation difficult. The disc so made was used for many lenses from 8″ F/2·9 to 36″ F/6·3.

4. The pencils of light are observed by means of a low-power microscope (1½″ O.G., N.A. 0·19, ×5 eyepiece, 160 mm. tube length) arranged to slide in a longitudinal direction along the lens axis and capable of movement in directions perpendicular to the lens axis. The whole of the apparatus, consisting of the lens, perforated plate and microscope, is mounted on an optical bench fitted with levelling screws so that good alignment can be obtained, and the pattern of images remains central as the microscope is racked in and out.

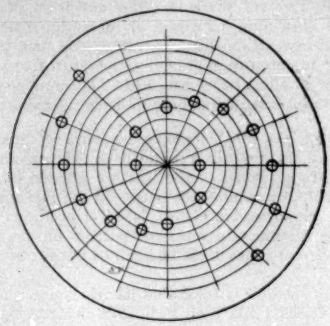


Figure 1. Zone plate.

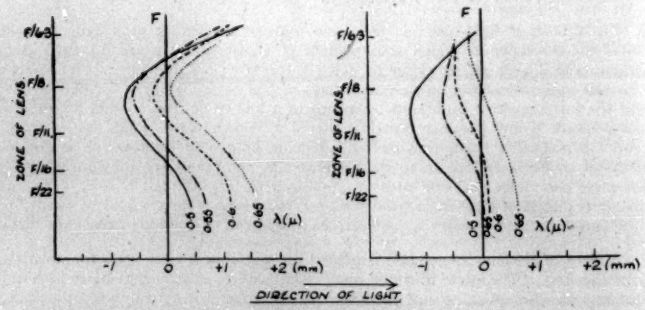


Figure 2. Spherical aberration: F indicates the position of the visual minimum fringe focus, using tungsten light and minus blue filter at full aperture.

On a stage placed at the focal plane of the microscope is mounted a graticule, consisting of a central dot and a number of concentric circles, such that, when viewed by the microscope, some six or eight concentric circles are seen.

5. Before commencing observation and measurement, the best position for observation of each pair of holes is found, and it is for this reason that a number of circles are necessary in the measuring graticule. On racking the microscope in and out it is found that at positions inside and outside the crossing point of the pencils, clear diffraction patterns having either a small concentration of light or a small dark spot surrounded by rings are visible. These are the most accurate points for measurement, and the nearest circle on

the graticule is chosen for the measurement. The process of measurement consists in moving the microscope until the pair of diffraction patterns corresponding to a given zone of the lens coincides with the particular circle inside and outside the crossing point of the pencils. Different circles of the graticule can be used for different zones, but the same circle is used both for the inside and for the outside pair corresponding to any one zone. The actual measurement consists in finding the distance moved by the microscope between the inside and outside coincidences, and it is thus clear that the crossing point of the pencils is half-way between these two points. It is convenient to use a vernier to read the movement of the microscope, and the zero of the scale can be at any convenient point.

The monochromator is set to give any colour for which the spherical aberration is to be measured, and by changing the colour after each determination the spherical aberration for a range of wave-lengths can be found. The curves shown in figure 2 were actually obtained with two lenses of 20″ focal length.

As a variation of this method, the microscope was fitted with a single cross-wire graticule and the actual distances between pairs were measured by moving the microscope in a direction perpendicular to the axis of the lens. Each pair of holes in the perforated plate must be brought into a convenient position for this measurement by rotating the plate in its holder. The results obtained by this method are equally accurate, provided the

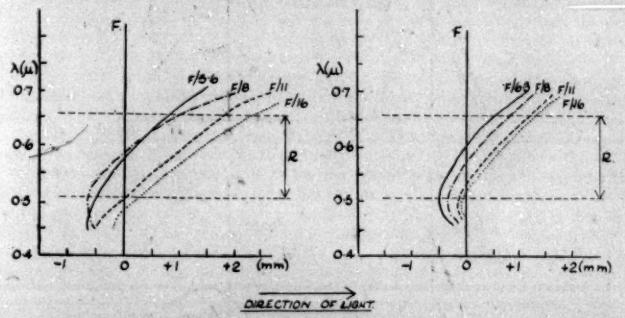


Figure 3. Chromatic aberration: F indicates the position of the visual minimum fringe focus, using tungsten light and minus blue filter at full aperture. R is the range of wave-lengths for which panchromatic film is sensitive with minus blue filter.

alignment of the whole set-up is accurate. By either method it was found that the readings obtained by a number of observers could be repeated within 0·1 mm., and the final curves were identical.

6. Before completing the test the perforated plate is removed and, by observing the monochromatic images of the star, the best positions of focus for each aperture and each colour are found. The series of curves obtained for the same two 20″ lenses is shown in figure 3. The white-light star focus is also found and the focus using any filter with which the lens may be used. To find the paraxial focus a knife-edge test gives the most accurate results.

7. The data obtained in these tests are most useful when used in correlation with a photographic test showing the maximum resolution of a test chart in relation to the visual focus setting. The photographs are obtained by moving the plate or film in small steps through the position of visual focus and taking a series of exposures of resolution charts of various contrast ranges. Such a series of curves is shown in figure 4. It will be seen that the depth of focus at each aperture and the shift of focus can be easily determined. The curves shown in figure 4 were obtained with white light, but similar curves with monochromatic light yield more information on the relation between the spherical and chromatic aberrations and the photographic image.

8. The chief disadvantage of the method of test lies in the decreasing accuracy obtained near the axis of the lens and the extreme difficulty (due to depth of focus) in finding the focus of the lens for very small apertures. The actual paraxial focus is only of theoretical interest and does not enter into these experimental methods ; moreover, at very small apertures, the depth of focus is so great that the position of the focus is unimportant from the photographic point of view. One very great advantage of the test lies in the ease with which the spherical aberration for any colour can be seen without the necessity of making any measurements. The relative movement of the diffraction patterns as the microscope is moved to and fro gives an immediate visual indication of the state of correction. Defects such as strain and bad figure are also immediately recognizable if the lens or perforated plate is rotated about the optical axis.

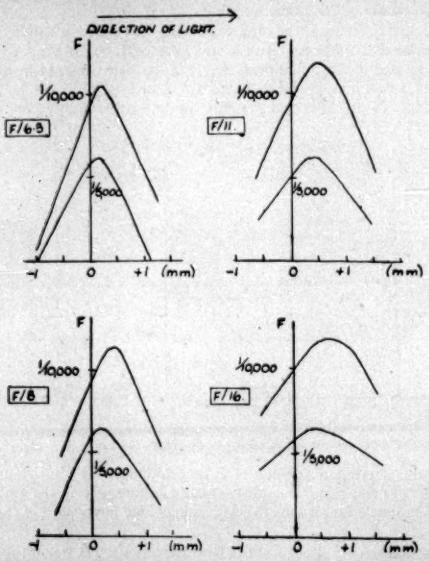


Figure 4. Resolving power: F indicates the position of the visual minimum fringe focus, using tungsten light and minus blue filter at full aperture. Upper curve, high contrast. Lower curve, medium contrast.

9. *Oblique tests.* A similar technique is applicable to the measurement of oblique aberrations, and particularly to the measurement of coma and astigmatism. The perforated plate used for such tests includes a central hole isolating a chief ray and pairs of holes in the meridian plane for the measurement of tangential coma. It is found most convenient to use a series of plates for different zones and to measure the sagittal rays for " D " pairs with different plates in which only the central and one " D " pair are perforated. A plate with concentric rings of holes corresponding to three zones of the lens, as described by Kingslake (1925), has been used with success, and the visual impression obtained is very illuminating. The analysis of the results is very similar to that described by Kingslake, and it is possible to obtain an indication of characteristic lines for the images. A further development of the method consists in placing the perforated plate in the plane of the diaphragm of the lens, and interesting results can be obtained by this method. The calculation of the results of observation is described fully by Kingslake. The measurements are best made by means of a stage micrometer fitted to the microscope.

10. *Astigmatism and curvature.* Instead of using the normal nodal slide and tangent bar method of obtaining the sagittal and tangential field curve, a very simple and rapid method can be used in which a series of horizontal and vertical lamps are placed at a distance and their image viewed on a screen with a travelling microscope. The movement of the screen towards or away from the lens for each part of the field is measured directly by means of a dial gauge fitted either to the lens carrier or to the screen carrier, and the field curves plotted directly. These lights also indicate very clearly the presence of coma and transverse chromatic aberration.

The astigmatism for different wave-lengths has been measured by using the normal nodal slide method and the vertical slit of the monochromator for the tangential focus. In order to obtain a horizontal slit for the sagittal lines, a right-angled prism is placed in front of the exit slit and rotated through 45 degrees when the slit appears horizontal.

In all such work the results are correlated with the photographic focus by taking photographs at small intervals through the position of focus at points off the axis of the lens, using charts having horizontal and vertical lines. Dial gauges are always used for measuring small movements. In this way the resolving power of the lens over the whole field for any position of the focal plane can be assessed.

11. *Distortion.* The distortion of camera lenses can be conveniently measured by an auto-collimation method. A graticule having a number of lines at measured distances is placed in the focal plane of the lens and one line is made to pass through the optical axis. The camera is then set up in a horizontal position with the entrance pupil of the lens as near as possible to the centre of rotation of a goniometer table, on which a good mirror is mounted vertically in front of the lens. The goniometer table is then rotated until the image of the centre graticule mark is made to coincide with the actual mark on the graticule. If the table is then turned until each successive graticule mark is made to coincide with its image, and the angle through which the mirror is turned is measured, the angular amount of the distortion is immediately found for each part of the field by taking the difference between the angle inside the camera given by the relation $d=f \tan A$ and the angle outside the camera as given by the rotation of the mirror A'. The value of f chosen may be that corresponding to small angles, but it is more convenient to use the value of f, which corresponds to minimum distortion at all parts of the field. The effect of moving the diaphragm is also found, and is a useful addition to the information obtained.

12. *Conclusion.* The theoretical aspect of the subject of aberrations and their measurement have been omitted, but much useful data can be obtained from the series of tests described. The accuracy of the results appears satisfactory, since repeated readings can be obtained by different observers and on different occasions, and there is no difficulty in arriving at the optical path differences and distribution of light in the image from a careful analysis of the results. Photographic results are very necessary and important, and give a very good indication of the effect of the various aberrations on the depth of focus and on the performance of the lens ; in fact, the two types of test are complementary and necessary.

REFERENCES

HARTMANN, 1904. *Z. InstrumKde.* **24**, 1, 33, 97.
KINGSLAKE, 1925. *Trans. Opt. Soc.* **27**, 221.

Mr. E. W. H. SELWYN. The performance of a lens changes in marked fashion according to the way in which it is used. It is common experience, for instance, that a photographic lens will give high resolution if the image is observed visually, and usually a much lower resolution if photographs are taken with it. Lenses frequently exhibit a fall in visual resolution, but an increase in photographic resolution on stopping down. In order to gain some insight into the reasons for this kind of behaviour the following argument has been used.

If we consider a test object composed of long parallel lines, the major effect of lack of sharpness in the image of the test object produced by any lens will be to decrease the contrast between the lines and the spaces between, and the diminution of contrast will be greater the closer the lines are together. If we use a test object in which the lines are not

sharp-edged, but such that the brightness of the test object across the lines can be represented by a sine curve (plus a uniform brightness, to avoid negative brightnesses), then the only effect of lack of sharpness is to diminish the contrast. The behaviour of the lens, at the point in its field where the test is made, can therefore be represented by a plot of the reduction of contrast against the wave-length of the sine curve or the separation of the lines or any such convenient variable. In theory (and except for the labour involved, in practice), this curve can be calculated from ray-tracing data or from interferometer patterns. It is, therefore, a possible representation of the purely optical properties of the lens.

Now the visual resolving power of the lens can be found from this curve if we know the minimum contrast which is discernible by the eye, under the conditions of observation. Experiment suggests that for the usual type of observation this threshold contrast is constant at about 3 per cent difference between the lightest and darkest parts of the image, whatever the actual separation between the lines observed, provided that the image presented to the eye, via the observing microscope, is of a comfortable size for viewing.

When the image is recorded on photographic material, the problem is more complicated. Nevertheless, it is possible to understand in a general way what happens. The turbidity of the emulsion layer will degrade the contrast of the image still further than the lens has already done, and the curve of contrast against line separation must be modified accordingly. Then allowance must be made for the change in contrast produced by developing the photographic image. This will give a new {contrast, line-separation} curve appropriate to the developed image. To estimate the resolving power we now need to know what the threshold contrast is. We do not know this accurately, but, so far as is known, the threshold contrast is greater with more grainy photographic material, and also greater for smaller detail, since the grain is relatively more important when small detail is being examined than when coarse detail is involved.

This way of approach to the problem of estimating lens performance has proved useful, and it is probable that when a description of the performance of a lens, which is fairly complete from the user's point of view, is required, the curve of contrast against line separation would be found a satisfactory means of description.

Mr. C. G. Wynne. It is of course true that to give a complete description of the performance of a lens would require curves (e.g. of the type described by Mr. Selwyn) for all parts of the image plane, all orientations of the (line) image, and all relevant wavelengths, for different positions of focus, together with photographic emulsion curves. A lens corrected to give the best possible result under one set of conditions (of object, light, emulsion, etc.) may not give the best results under different conditions ; but these effects, important as they are for special cases, are small compared with the differences in performance between well corrected and poorly corrected lenses, and as Prof. Martin said at a previous discussion, " generally .... the best corrected lens will give the most readily interpretable image ". It should not be forgotten that the user of a photographic lens normally requires it to give satisfactory results under a wide variety of conditions and over the whole angular field for which it is designed, being unconcerned with localized or specialized excellences, so that for such purposes a single index, or simple set of indices, of lens performance under defined average conditions might be devised. If the work Mr. Selwyn has described were to lead, *inter alia*, to the definition and general adoption of such index or indices, I think it would be most valuable. It would necessarily be empirical, approximate and liable to mislead under extreme conditions of use (as is, for example, a figure of photographic emulsion speed), but at present the lens user has no kind of indication of quality save price and name, and the designer no general comparative standard of performance, a position which is surely not conducive to progress.

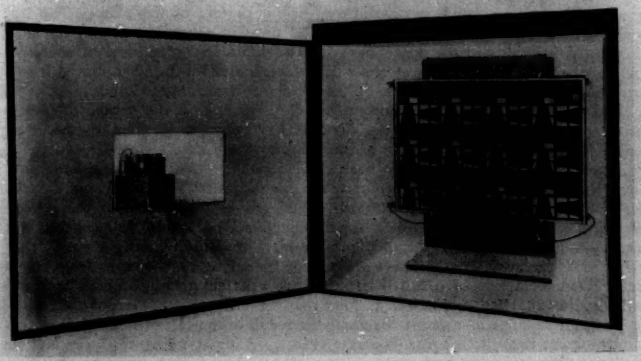
Mr. H. K. Bourne. When camera lenses are tested by photographing a resolution chart, the problem arises of obtaining an even illumination and of avoiding specular reflection over the area of the chart. This problem has been overcome in a simple and effective manner by making use of the principle of the light integrator in order to obtain a uniformly and diffusely lighted surface.

Resolution test charts are often printed with a glossy surface, so that it is not easy to avoid specular reflection. When this does occur, the results produced are misleading,

since parts of the image may be over-exposed and fine detail completely lost. It is also important that the whole of the area of the chart should be illuminated uniformly, so that any variation in illumination produced by the lens over its field may be measured.

In this apparatus the resolution chart is placed inside a wooden box painted on the inside with zinc oxide paint. The chart, which is situated near the back of the box, is illuminated by two experimental Mazda Type MCF tubular fluorescent lamps, each rated at 40 watts. These lamps are placed behind the board on which the chart is mounted and the illumination on the front of the board is obtained indirectly by reflection from the sides and the front of the box. The photographs are taken through a rectangular aperture cut in the door of the box, the camera being supported on a bracket mounted on this door. The size of this aperture is just sufficient to cover the angle of view of the camera.

The apparatus is shown in the accompanying figure. The box is a cube of 2′ 6″ size. Two special fluorescent lamps, only 24″ in length, have been used, so that these could be accommodated inside the box. Since these lamps are situated behind the definition chart, they are shielded from the view of the camera so that no direct illumination from the light sources falls on the camera lens. The illumination produced on the test chart is approximately 40 foot-candles, and this is practically uniform in intensity over the whole



of the area of the chart. A particular resolution chart which has been used is printed with a glossy ink on cartridge paper and the surface of this appears to be quite matt when it is placed inside the box.

The tubular fluorescent lamps are used partly because they are excellent diffuse sources, and partly because they give a colour rendering approximating closely to daylight. In many cases the lenses being tested will be required to work under daylight illumination. If it were necessary to carry out tests with any particular colour of light, then of course, this box could be illuminated with discharge lamps giving the appropriate colour. The tubular fluorescent lamps are convenient to use, as they may be switched on or off as desired without any appreciable delay. Owing to the high efficiency of these lamps, a number of them may be used in a bank to obtain a very high illumination with little heating, if this is required.

Since the residual aberrations of a well corrected photographic lens are so small that loss of definition in the recorded image is due to the film emulsion rather than to errors in the lens, the photographic technique employed in carrying out tests on lenses is of great importance.

It is necessary to choose an emulsion and developer to minimize the effect of grain. A thin emulsion should be used in order to reduce the effect of scattering and diffusion. Emulsions such as Microfile or Microneg are quite satisfactory for testing the definition

of lenses. Where even finer resolution is required, Kodak maximum resolution plates or other similar materials are quite suitable.

[A number of enlargements of negatives using various emulsions with different developers were shown to demonstrate these points. Also a number of photographs of the definition chart taken with the Leica and Contax cameras using large aperture lenses were displayed.]

Mr. J. Hasselkus, C.B.E. For a great many years we used to test our photographic lenses by taking photographs of a test chart arranged on a board $8' \times 10'$ and having on it printed matter of various sizes—circles, squares, horizontal lines and vertical lines—arranged from the centre outwards fairly symmetrically. Photographs were taken from distances up to 30 feet, according to the length of the focus of the lens under test, and also having regard to the angle which the lens had to cover. The quality of the lens was judged by the definition all over the plate after focusing for the best position in the centre.

This method proved very satisfactory up to the time when we designed our F 4·0 Wide Angle Lens, which was constructed to cover a square plate, the side of which was equal to the focal length of the lens. This lens was specially designed for aerial survey work and was largely used in America and in Canada. In fact, the $8\frac{1}{2}''$ focus lens became the standard for this kind of work in those two countries.

Each lens which was sent to America underwent a test at the Bureau of Standards, and the test they applied was a resolution test ascertaining the number of lines per millimetre which the lens would resolve in the various regions of the plate. The test object consisted of batches of vertical and horizontal lines in separate squares, the black lines being equal in width to the space between two lines. A good deal of correspondence took place with the users of these lenses in America, and we found it advisable to arrange our testing methods on similar lines. This procedure was adopted about eight or ten years ago, and we must confess that it has proved very valuable in the testing of photographic lenses, and was still in use when the Air Ministry began to wonder whether the photographic lenses supplied by makers in this country were best adapted for their purpose, and a great deal of work has since been carried out by Messrs. Kodak, Ltd., and others to arrive at a method to give an absolutely reliable indication of the value of any particular photographic lens for aircraft photography. It was at this time that the question was debated as to whether the best object for these lenses should be of low contrast or of high contrast, and I think opinions regarding this subject are still divided.

If you desire to obtain the finest instrument for taking photographs from the air, it is surely reasonable to suppose that if a lens will give a perfect image of a high-contrast object, it will also do so of a low-contrast object, but that from the point of view of the lens designer it is very much more convenient to examine the result of the test if it is made with a high-contrast object than with a low-contrast object. A great many tests carried out on a number of different lenses have always shown that a lens which appears better when tested with a high-contrast object also shows up better when a low contrast object is used.

The test object mentioned above occasionally led to spurious resolution results, showing in some cases that finer lines were resolved, whereas coarser lines were not resolved. This particular drawback of this object was successfully overcome by the Cobb Chart, wherein the test object consists of two short lines; the thickness of each is equal to the space between them, and they are repeated some distance away. The width of these lines can be arranged to provide a suitable gradation from coarse to fine. This test object, I believe, is now almost universally used by makers of photographic lenses. It is only necessary to arrange the size of the test object to suit the focal length of the objective if a standard collimator is used in the test. By a careful study of a photograph taken of the Cobb Chart it is quite easy to plot the resolution over the whole field, both for vertical and horizontal lines, if test photographs are taken at stated intervals, say every 2 or 3 degrees, according to the focal length of the lens.

We occasionally have our lenses tested by the National Physical Laboratory on the Hilger Interferometer, which also gives very valuable information on the general behaviour of the lens. I must admit, however, that we find it difficult to lay our finger on the culprit which is causing the falling off of the definition, but the interferograms are a very good guide to the uniformity with which the manufacture is being carried out.

I should like to draw attention to a method of testing photographic lenses described by Wetthauer, of the Charlottenburg Reichsanstalt, in 1921. His method is to arrange the

objective in front of a collimator and take photographs on a plate which makes an angle of 10° with the horizontal, the test object in the collimator consisting of two bright lines separated by a space of equal width to the white lines drawn vertically through the centre of the test object, and on one side of these two lines a series of horizontal lines, slightly thicker than the centre lines, and on the other side a series of horizontal lines similar in thickness to the centre lines. The width of the vertical lines should be about 1/100 of the focal length of the objective to be tested. In the photograph the vertical lines appear as two streaks of light spreading out from a very sharply defined constriction, and the horizontal lines appear as a few short horizontal lines well defined and deteriorating into a smudge below and above. This method can be used with monochromatic light of different wave-lengths. Respective photographs will then show the amount of chromatic aberration present as the position of the constrictions of these lines will vary along the plate. The determination of the spherical aberration can be carried out in the same manner by making various zones of the objective active by stopping down the remainder. Astigmatism and roundness of field are clearly indicated by the position of the smallest constriction of the vertical lines, and the best definition for the horizontal lines. This is obtained by swivelling the lens and plate through definite angles. This method appears to be eminently suitable for examining prototype lenses, but it would appear to be rather cumbersome for a routine test of lenses. We ourselves have not used this method, but I would be interested to know whether anyone else has made use of it in this country.

At present the design and computation of photographic lenses consists of a process of balancing the inevitable residual errors against the ten conditions which should all be fulfilled in an ideal photographic lens, and at present we have to rely on the judgment and experience of the designer. If a method could be devised whereby the errors against these conditions could be separately ascertained, it would help designers materially. All these errors should be measured through a prescribed region of the spectrum, and all those caused by oblique pencils should be measured at stated angles in addition. A comparison of results given by a reasonably good lens and by an excellent lens of comparable type should give the designer a great deal of useful information. I admit that a good many of the errors can already be measured by known methods, but not all.

Mr. C. A. PADGHAM. There is one point on which I disagree with Mr. Hasselkus. Photographic resolution results obtained with high-contrast test objects will not necessarily indicate what will happen with a low-contrast test object. In the latter case, the aberration " fringe " around the image plays a more important part than in the high-contrast image, in which the " fringe " is rendered less effective by the high central intensity of the image. Therefore, a lens should always be tested with low-contrast charts if it is subsequently to be used for low-contrast work, even though the high-contrast results may be more valuable in design.

Prof. L. C. MARTIN. The discussion this afternoon has recalled to my mind a vague impression derived from previous experience that many people who are interested in the production of lenses and in their improvement are, for some reason or other, sceptical of the usefulness of the actual measurement of their defects. It is obvious that such a procedure has an application in which an individual system (say a large astronomical mirror) is to be figured and corrected; but in cases of systems in large scale production the tendency would generally be to rest content with a functional test, e.g. photographic resolving power, etc.

Thus in spite of the early development of the Foucault test and various other schemes from Hartmann, Chalmers, and more modern followers, and especially in spite of the quickness and accuracy available in the interferometer tests, interest in this field of work has languished, except in a few laboratories with special interests.

In some cases at least, factories have found that, in spite of their best efforts, lenses could not be expected to conform exactly to the predictions of design. Some adjustment of separations, etc., might be required in order to compensate for variations in the refractive index of the glass, even if the radii were made to conform accurately to specification. If no particular individual is typical, it would certainly be natural to pay attention simply to the functional test.

However, in view of modern improvements in annealing it should be possible to make individual systems approximate more and more closely to design; and, in fact, the standard of reproduction is often extraordinarily high. The analytical and precise test should then come into service; it should be carried out sufficiently thoroughly for the true features of the actual performance in the functional test to be understood. It can be stated that such an understanding is still lacking in the case of many photographic and microscopic objectives.

Such a deeper understanding of performance is one line of approach toward the improvement of lenses. It implies also some progress toward the fuller analysis of experimentally determined aberrations. The study of the relative incidence of the higher aberrations in different designs, and of the balance of primary against higher aberrations, offers a large field for work of a practical character. It might be said that this would best be done by ray tracing; but, in fact, the time absorbed in tracing skew rays in sufficient numbers to make possible a satisfactory analysis of oblique aberrations would be very great.

The precision of practical tests should be adequate, especially if refinements like those of Platzeck and Gaviola or the careful use of the interferometer are remembered. We seem to have reached the limits which can be obtained more or less readily on empirical lines, and the attainment of the further levels of performance that may be reached when the use of aspheric surfaces is better developed will require a more intimate study of the art of design based on a truer estimate of the relations between aberration and performance.

Mr. H. H. Hopkins. The word "aberration" suggests, in connection with the subject of this discussion, the need for a distinction between the aberrations of a lens system as such and their effect on the performance of the system under any given conditions. For this latter will, in addition, depend upon the characteristics of the object (mean intensity, size of detail, contrast, etc.) and upon those of the receiving screen (threshold, resolving power, contrast sensitivity, etc.). It will also be greatly affected by the choice of focus. A study of the experimental results obtained in lens-performance tests—e.g. photographic resolution—reveals the inadequacy of existing lens calculations in relation to the problem.

It is now more than a century since Airy's results were published, yet it seems that the majority of lens designers still regard the ray theory as adequate. Even more surprising is the attempt to look upon "aberration" and "diffraction" as two separate agencies tending to "impair definition". That, in the presence of large amounts of aberration, the ray theory is capable of approximately describing the limits within which light is seen is indubitably true. Where it fails is in providing a description of the distributions of light in the image when viewed under different conditions of focus. Fundamentally, the error seems to lie in assuming that because differences in the ray-intersection distances are measured with reference to the parallel ray focus, they are, therefore, intimately related to the image. In fact they are an arbitrary (and, in my opinion, unfortunate) parameter describing the departure from sphericity of the wave-front in terms of the departure from homocentricity of the range of wave-normals associated with it. At best, the angular ray aberration denotes the rate of change of such wave-front departure with aperture; and, without assuming a given functional relationship between the two, will not give the departure itself except by numerical integration using a number of ray traces. If the wave-front is perfectly spherical the ray trace and optical path calculations will agree; but only when this is true or very nearly so. Further, given the form of wave-front, the range of rays is uniquely determined: the converse is not true, except—as suggested—when a given functional relationship is assumed between the two.

Bearing in mind the fact that lens imagery is more accurately regarded as a wave problem, and noting the behaviour of the wave-equation near a source or focus, the solution falls mathematically into two parts: the one, the Huygens envelope solution, valid outside the region bounded by a small sphere surrounding the centre to which the wave proceeds, and an integral solution valid within this region. In arriving at the final form of the wave-front, the ray concept has its rightful place in determining the paths that different parts of the wave-front take through the system. It is, in fact, the variation of such paths

due to the presence of aberration that militates against an aberration theory related to the axis and valid for higher apertures and larger fields, for it requires the aberration contribution at a given surface to be a function of those due to the preceding surfaces.

The second part of the solution is one in which a good deal of investigation still requires to be undertaken. Nevertheless, the work that has been done is very often ignored. Airy, Lommel, Struve, Strehl and Rayleigh are a few of the names associated with fundamental advances in the field of the diffraction effects near the focus of an approximately spherical converging wave. More recently Steward * has dealt with the effects associated with the Seidel first-order aberrations. One defect of his results lies in the unsuitability of the series for cases of aberration and differences of focus amounting to a few wavelengths—a defect the surmounting of which has been attempted for axial images in work which the writer hopes to publish shortly. The link between the lens as designed and its performance as produced is to be found in the light-distribution curves associated with the given wave-front when viewed under different conditions of focus. Because of this, when a suitable theoretical account of these distributions is available, the Camera Lens-testing Interferometer will become an immeasurably more useful instrument.

There are two further questions deserving attention. The first concerns a theoretical result which supports the use of projection methods for investigating experimentally the form of light distributions associated with different conditions of focus for a given lens. It can be shown for moderate angular apertures and fields that, between conjugates of magnification m, the form of light distribution in an enlarged image is the same as that in a reduced image between these conjugates, but on a scale m times as large. A change in focus δx for the reduced image is equivalent to a change $m^2\delta x$ for the enlarged image. Both these results lead one to expect a considerable simplification in experimental technique.

Consider a point $P'(\rho', \phi')$ in a plane of focus δx from the paraxial plane, the co-ordinates being with reference to the intersection of the principal ray with this plane. If η is the distance of the Gaussian image on the paraxial plane from the axis of the system, the co-ordinates of P′ are

$$x' = -\delta x,$$

$$y' = \eta\left(1 + \frac{\delta x}{F}\right) + \rho' \cos\phi',$$

$$z' = \rho' \sin\phi',$$

where $(x\,y\,z)$ is a rectangular system with origin at the Gaussian axial focus, and chosen so that the extra-axial image lies in the plane $z=0$. The co-ordinates of a point P in the exit pupil are $(F,\ \rho\cos\phi,\ \rho\sin\phi)$, F being the distance of the node from the paraxial image plane. † The length PP′ is thus given by

$$d = \text{const.} - \left\{\frac{\rho\rho'}{R}\cos(\phi-\phi') + \delta x\,\frac{\rho^2}{2RF} - \delta x\,.\,\frac{\eta^2}{RF}.\,\rho\cos\phi - \frac{\eta\rho'}{R}\cos\phi' + \delta w\right\},$$

where δw is the advancement of the wave emerging through P with reference to the "paraxial sphere" centred on the Gaussian image, and R is the radius of this sphere. Thus the intensity at P′ can be written as the squared modulus of the function,

$$E = e^{-\frac{i2\pi}{\lambda}\cdot\frac{\eta\rho'}{R}\cos\phi'}\int\rho d\rho\int e^{\frac{i.2\pi x}{\lambda}\left[\frac{\rho\rho'}{R}\cos(\phi-\phi')+\delta x\,\frac{\rho^2}{2RF}-\delta x\,\frac{\eta}{RF}\rho\cos\phi+\delta w\right]}d\phi.$$

Along a given ray-path (ρ=const.), δw is the same for both the "reduction" and "enlargement"; $\eta_2 = m\eta_1$; $R_2 = mR_1$; $F_2 = mF_1$, where suffixes $_2$ and $_1$ refer to the reduced and enlarged images respectively. Thus, if one writes

$$\rho_2' = m\rho_1',$$

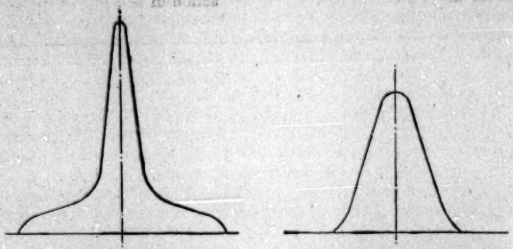
$$x\delta_2 = m^2\delta x_1,$$

* Steward, "Aberrational Diffraction Effects", *Philos. Trans.* A, **225**.

† This assumes the stop position as identical with the Gaussian nodal plane—an assumption valid in most photographic objectives.

the expressions under the integrals are identical, which demonstrates the proposition to the accuracy contemplated, the degree of which will be evident.

The rough forms of light distributions for axial images in the presence of aberration under different conditions of focus are easily seen with a star imaged by a photographic lens. They are equally appreciated if a transparent line on an opaque graticule be illuminated and projected on a distance screen. Curves of the form A and B (see figure) can be obtained by focusing for " sharp central maximum " and for " minimum fringe ". They might be called the " sharp maximum " and " minimum fringe " conditions of



focus. That under certain conditions the former focus gives better photographic results on low-contrast test objects and the latter on high-contrast test objects is at once apparent from the curves shown by Mr. Dickson. A parallel to these results is to be found in the well-known difference between the focus required for best visual resolution and " definition ", which latter seems usually to mean what has been denoted here by minimum fringe. It is immediately seen how the two forms of light distribution affect Mr. Selwyn's {contrast, log R} curves for the lens, and he is to be congratulated on providing such a simple means of presenting the problem.

Mr. J. W. Perry. The optics of the burning glass and of the lens magnifier may be traced back to early times, but what in a theoretical sense may be called higher optics, based upon an analytical approach to the consideration of lens action, embraces, from the time of Euler, a period of scarcely two hundred years. In the course of this, opticians have gradually become familiar with the nature and analytical form of the defects of lens systems; but as the performance of a lens depends ultimately upon the formation of an image upon some photo-sensitive surface—be it the retina or a surface ultimately to be projected upon the retina—there has remained scope for appreciable latitude of judgment, particularly as in actual cases confusion of thought is often occasioned by different standards of judgment being employed by different authorities. One cannot rightly distinguish in this matter between optical and other standards and scales of performance, for the whole matter of performance as such in this connection is inherently optical, bound up, though it may nevertheless be, with phenomena and processes subsidiary and contributory to the main purpose of any optical apparatus as an aid to seeing. It appears, however, to be very much less difficult to prescribe the function of a lens and to test for functional errors than it is to measure and standardize the gross performance in a manner which interests the user.

Waves of light are the essential basis of the phenomena here involved, and provide the vehicle for the propagation of energy, which may be made to perform useful work in various ways, subject always, of course, to certain fixed limitations of a known character. We are in this present connection concerned with systems, mainly photographic, which produce visual images, and we cannot therefore divorce attention entirely from visual processes ; for it would be distinctly misleading, in an attempt to rule out visual factors involved in the interpretation of photographic images, if, for example, one were to apply

in the examination of an image such a measure of supplementary optical aid as would make interpretation in the usual sense impossible. This is, of course, ultimately a matter for readjustment of point of view, but " interpretation " in terms of statistical analyses of microphotometric observations would not be regarded by a normal user as synonymous with interpretation in the usual sense. Accepting, however, for the time being the exclusion of visual phenomena from our considerations of optical performance, it is clear that tests of performance should, in order to have constructive value, be functional tests, and should embrace the following :

(*a*) A test for the degree of perfection of lens action in terms of light waves, by which one may judge, against the standard of the physical concept of the ideal lens, the capacity of any given objective to redirect light waves for the purpose of forming images.

(*b*) A test to assess performance under given conditions of use, by which one may determine the actual degree of correlation attained between object and image.

Considering these two aspects of lens testing, one may note an important distinction between them. In the first case, the ideal, in a practical sense, and physical standard by which any lens is judged, and also the degree of approach to such an ideal, are definite and quantitative. The methods, involving the now well-known interferometric tests, are obvious and readily applicable and yield results enabling the measure of approach to the standard to be precisely stated and quantitatively analysed. In the second case nothing is definite. An unlimited correlation between object and image is no doubt desired, and to some extent, depending upon the material outlay allowed, it is unlimited ; but quite apart from the limitations imposed by the finiteness of light waves, unlimited correlation is physically meaningless on account of an ultimate inherent vagueness or indefinition in the structure of the object.

If we may exclude qualitative properties such as colour from present consideration, and confine attention to two-dimensional objects, it will be seen that a measure of correlation between object and image demands first an analysis of the photometric structure of the object, assuming this to be built up from cells or surface elements, finite but negligible in size, and uniform in brightness. In order not to invalidate the application of this analysis, a similar assumption needs to be made in relation to the structure and corresponding surface elements of the image. Thus, in the absolute sense, the scale of correlation is indefinite. Scales of correlation may, however, be constructed relative to the physical limitations of specific imaging systems. This involves incidentally a clear discrimination of resolving power and defining power,* a subject which was considered earlier by Strehl, Michelson and others, but is still in an unresolved condition. A further development of this and allied subjects will be given in a future paper. Suffice it here to say that the complexity of image formation and the need for tests to have a constructive value, and not merely to be criteria of acceptance or rejection, emphasize the need for tests of a functional character.

The problems of photographic lens testing may thus be considered as linked severally with the three main subdivisions or stages in the process involved in the use of such systems, viz., Lens Action ; Image Formation ; Interpretation, the latter involving problems in which visual contrast theory, microphotometry and mathematical analysis may each play an important part. Each such stage is causally connected with the preceding through the contributory phenomena employed.

It is clear that no one test can convey information which will completely determine the performance of a photographic objective in this way in relation to given conditions, which may influence performance radically by their general selectiveness. For practical purposes it is, of course, very desirable that tests, while comprehensive, should also be simple in their procedure and informative in their results, and in view of the causal connection between the respective stages of the process, the only way in which the desired condensed informativeness can be obtained in such circumstances is, in the writer's opinion, provided by an interferometric test, giving information in terms of wave deformation. This method

* The burning glass, or its modern variant the infra-red spectrometer, provide incidentally an example of how defining power might be measured on a thermal basis.

has the desired qualities above referred to, and also has advantages of a theoretical nature in that the monochromatic aberrations of any given system can be measured in the same terms as those used by the lens designer.

The advantage of a method for ascertaining the errors of compliance with the conditions of photographic lens design has been referred to by Mr. Hasselkus on previous occasions, and provided at one time a problem which it was agreed was then satisfactorily solved.* The method adopted was based upon the use of the interferometer. From an interferometer test one may, indeed, derive information for lens designers which will very considerably supplement the knowledge they already have on the errors in the performance of any given system. It would be interesting to know if Mr. Hasselkus refers in his present remarks to analytical errors and conditions of this kind, or to defects of gross performance under specific conditions involving integrative processes in their analytical specification.

* Hasselkus, 1923/1924. *Trans. Opt. Soc.* **25**, 97.

REVIEWS OF BOOKS

Proceedings of a Conference on the Ultra-fine Structure of Coals and Cokes by The British Coal Utilization Research Association. Pp. 366. (London: H. K. Lewis and Co., Ltd., 1944.) 25*s.* net.

Scientific research is one thing; *industrial* scientific research is, more often than not, quite another, for it is prosecuted, quite legitimately, with a view to increased financial profits. Its results, if possessing financial potentialities, are, customarily, carefully safeguarded; such results as are freely published are usually of little or no pecuniary value, but possess very considerable value for propaganda purposes. For industry has discovered that its status, in the eyes of an unintelligent, unscientific public, is much improved if it is shouted from the house-tops, or even merely whispered, that it is prosecuting research. Present indications are that this condition of affairs will be aggravated in the post-war years.

Here are 22 papers devoted to matters connected with the structure of coal (not necessarily the *ultra-fine* structure of coal as understood by physicists), read at a conference held at the Royal Institution on 24 and 25 June 1943. There are also a foreword contributed by Sir Evan Williams, President of B.C.U.R.A., and an introductory address by J. G. Bennett, Director of Research, B.C.U.R.A., and arch-propagandist of the notion that the energy of coal is most effectively and scientifically realized by consumption of the fuel in open fire-grates!

The papers are of very unequal merit: none is of outstanding quality; none, so far as I can see, after a careful reading of the whole volume, will contribute one single item to a possible improvement in the technique of coal utilization. There are papers devoted to the investigation of the internal structure of coal from determinations of the heat of wetting of coals, from optical, elastic, magnetic, electric and x-ray measurements. Most of the work is of the kind the late Professor H. E. Armstrong would describe as "unripe fruit". And this is not to be wondered at. For coal substance is not elemental in character: it is a conglomerate of many substances. At present coal science has not passed out of the stage of being purely a descriptive science, like biology and geology. One thing is certain: if coal science is ever to achieve the stature of an *exact* science that end can only be reached by the application of modern statistical theory to ascertained results. One fact about coal is crystal clear: it is composed, at least in part, of one or more of four main rock types, designated by Stopes as *vitrain*, *clarain*, *durain* and *fusain*. Modern refinements of this classification refer to fusinite, micrinite, vitrinite, exinite and what not. Surely, if any ascertained physical characteristic of a coal is to have any significance whatsoever, this can only refer to the ascertained proportions in which these constituents are present. And this at present is not done. No wonder Dr. Stopes remarked at the Conference (p. 350): "I feel I must emphasize one elementary fact that most of the contributors to this Conference seem to have forgotten, namely, the

heterogeneity of coal. If I am told that these ' micellae ', these ' capillaries ', these ' crystalline ' things are all identical, and they are just ' coal ', I say look again down your microscope ", and that R. A. Mott should have stated (p. 159) that " the descriptive tags of the coals would be better omitted, and certainly the minds of old-fashioned, as well as of the new, exponents of coal research should not be confused by a terminology which is novel, maybe misleading, and certainly without merit ". To sum up my lasting impression after toiling laboriously through this book, I would say that present-day exponents of coal research reveal too little acquaintance with the canons of inductive logic (space fails me to give numerous examples of this), and that the science itself, as here presented, requires a vigorous application of Occam's razor. J. S. G. THOMAS.

A German Physics Reader, by J. E. Calthrop. Pp. 83. (London: William Heinemann, 1943.) 7*s.* 6*d.* net.

This little reader is intended primarily for the use of studcnts taking an examination in which a knowledge of German as part of the equipment necessary to a physicist is tested. It does not purport to teach the language to the beginner, who, indeed, without any grammatical knowledge, would not be able to profit from the book at all. There are no notes, even on interesting or idiomatic points of grammar which present themselves in the extracts given. The sole contents are a number of passages in German, with translations of the earlier ones into English, and with vocabularies for the latter ones. There can be no two opinions on the excellence of the selection made. None of the passages is dull reading, many of them are on topical aspects of physics, and the historical ones really throw light on their subjects.

The translations given for the earlier passages are very well done, judged by the standard which a professional translator would adopt. That is, they read smoothly in English, and convey the sense of the original. This very excellence may well be a difficulty to the less advanced students, since it implies that the translation is far from literal ; the reader cannot, even by counting up commas and full stops, identify the German phrase which corresponds to the English one, for dependent clauses in the one language may become independent sentences in the other, or may not be represented except by an adjective in a preceding or following phrase.

After page 25 no translation is given, but only a selected vocabulary, confined strictly to the technical words ; such an idiom as *Zustande kommen* is introduced in the text without reference to it in the vocabulary. The book concludes with a selection of passages actually set in public examinations, and here no vocabulary or other help is given.

Within its aim, therefore, the book is a very good one indeed, but it will be more useful when used in class by those fortunate enough to have a teacher than when used by the private student. The latter will suffer by the rather distressing frequency of misprints, due no doubt to war conditions. It is hardly an exaggeration to say that there is some error in practically every mathematical formula. Among the few other blemishes which have been noticed are the translation on p. 31 of " Krümmung " as " curve ", where " curvature " would fit the context better, and of " Verstarkerröhre " as " valve amplifier " instead of " amplifier valve ". Is not " Aluminiumblech " sheet rather than plate, and in an article on optical gratings, is " order " sufficient as a translation of *Grössenordnung* ? It might easily be taken by the student to refer to the order of a spectrum and not to an order of magnitude. J. H. A.

Experiment and Theory in Physics, by MAX BORN. Pp. 44. (Cambridge: University Press, 1943.) 2*s.* net.

This book is an expanded form of an address given to the Durham Philosophical Society and the Pure Science Society, King's College, at Newcastle-upon-Tyne in May 1943, and it owes its origin, Professor Born tells us, to his conviction that the philosophical ideas of Eddington and Milne, according to which the laws of nature are manifest, without appeal to experiment, to the mind well trained in mathematics and epistemology, are a considerable danger to the sound development of science. He has accordingly attempted to trace the mutual relationship between theory and experiment

in the actual historical development of science, with special emphasis on modern progress, of which he has had first-hand experience, and which at first sight appears to offer the most plausible evidence for the Eddington-Milne view. His conclusion is that in every instance theory and experiment interact with one another, and both are necessary to the final result. "We are in a jungle, and find our way by trial and error, building our road *behind* us as we proceed." Experiment alone can indicate the direction we must take, and theory alone can make that direction visible as a road along which advances can be made. Professor Born's treatment of modern ideas is most illuminating, and he makes clear beyond all doubt that experiment has made an essential contribution to them. It may be argued that the problem still remains whether the necessity for experiment arises from the imperfection of our minds or belongs to the nature of things, for presumably neither Eddington nor Milne would deny that, as a simple matter of fact, we have attained our present knowledge of laws of nature through experience. On this point the reader must form his own judgment, but he is left in no doubt concerning Professor Born's view of the matter. It would be a mistake to attempt to summarize the lecture, for it is itself a highly condensed summary, and should be read by all who are interested in the broader implications of physical theory. The eminence of the author is a sufficient indication that his views are entitled to the greatest respect, and the manner of his presentation of them makes the reading of the book a pleasure as well as an obligation. H. D.

RECENT REPORTS AND CATALOGUES

Mechanical Properties of Metals and Alloys, by J. L. Everhart, W. E. Lindlief, J. Kanegis, Pearl G. Weissler and Frieda Siegel. (National Bureau of Standards Circular C447, December 1943.) Pp. 481. U.S. Department of Commerce, Washington, D.C. $1.50.

Determination of Particle Size in Sub-Sieve Range. (Report of Discussions.) Pp. 69. The British Colliery Owners Research Association, Aldwych, London W.C. 2, and The British Coal Utilisation Association, Rickett Street, London S.W. 6.

Science in the Universities. (Report of the Association of Scientific Workers to the University Grants Committee of the Treasury, March 1944.) Pp. 44. Association of Scientific Workers, 73 High Holborn, London W.C. 1. 1*s*.

The Teaching of Mathematics to Physicists. (Report of the Mathematical Association and the Institute of Physics, 1944.) Pp. 20. The Institute of Physics, The University, Reading, Berks.